Don Harris (Ed.)

LNAI 4562

Engineering Psychology and Cognitive Ergonomics

7th International Conference, EPCE 2007
Held as Part of HCI International 2007
Beijing, China, July 2007, Proceedings

HCI 2007 INTERNATIONAL

Springer

VISIT…

LANZAROTE Caliente.com

Lecture Notes in Artificial Intelligence 4562

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Don Harris (Ed.)

Engineering Psychology and Cognitive Ergonomics

7th International Conference, EPCE 2007
Held as Part of HCI International 2007
Beijing, China, July 22-27, 2007
Proceedings

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editor

Don Harris
Cranfield University
School of Engineering
Department of Human Factors
Cranfield, Bedford MK43 0AL, United Kingdom
E-mail: d.harris@cranfield.ac.de

Library of Congress Control Number: Applied for

CR Subject Classification (1998): I.2.0, I.2, H.5, H.1.2, H.3, H.4.2, I.6, J.2-3

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-540-73330-2 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-73330-0 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12082889 06/3180 5 4 3 2 1 0

Foreword

The 12th International Conference on Human-Computer Interaction, HCI International 2007, was held in Beijing, P.R. China, 22-27 July 2007, jointly with the Symposium on Human Interface (Japan) 2007, the 7th International Conference on Engineering Psychology and Cognitive Ergonomics, the 4th International Conference on Universal Access in Human-Computer Interaction, the 2nd International Conference on Virtual Reality, the 2nd International Conference on Usability and Internationalization, the 2nd International Conference on Online Communities and Social Computing, the 3rd International Conference on Augmented Cognition, and the 1st International Conference on Digital Human Modeling.

A total of 3403 individuals from academia, research institutes, industry and governmental agencies from 76 countries submitted contributions, and 1681 papers, judged to be of high scientific quality, were included in the program. These papers address the latest research and development efforts and highlight the human aspects of design and use of computing systems. The papers accepted for presentation thoroughly cover the entire field of Human-Computer Interaction, addressing major advances in knowledge and effective use of computers in a variety of application areas.

This volume, edited by Don Harris, contains papers in the thematic area of Engineering Psychology and Cognitive Ergonomics, addressing the following major topics:

- Cognitive and Affective Issues in User Interface Design
- Cognitive Workload and Human Performance
- Cognitive Modeling and Measuring
- Safety Critical Applications and Systems

The remaining volumes of the HCI International 2007 proceedings are:

- Volume 1, LNCS 4550, Interaction Design and Usability, edited by Julie A. Jacko
- Volume 2, LNCS 4551, Interaction Platforms and Techniques, edited by Julie A. Jacko
- Volume 3, LNCS 4552, HCI Intelligent Multimodal Interaction Environments, edited by Julie A. Jacko
- Volume 4, LNCS 4553, HCI Applications and Services, edited by Julie A. Jacko
- Volume 5, LNCS 4554, Coping with Diversity in Universal Access, edited by Constantine Stephanidis
- Volume 6, LNCS 4555, Universal Access to Ambient Interaction, edited by Constantine Stephanidis
- Volume 7, LNCS 4556, Universal Access to Applications and Services, edited by Constantine Stephanidis
- Volume 8, LNCS 4557, Methods, Techniques and Tools in Information Design, edited by Michael J. Smith and Gavriel Salvendy

- Volume 9, LNCS 4558, Interacting in Information Environments, edited by Michael J. Smith and Gavriel Salvendy
- Volume 10, LNCS 4559, HCI and Culture, edited by Nuray Aykin
- Volume 11, LNCS 4560, Global and Local User Interfaces, edited by Nuray Aykin
- Volume 12, LNCS 4561, Digital Human Modeling, edited by Vincent G. Duffy
- Volume 14, LNCS 4563, Virtual Reality, edited by Randall Shumaker
- Volume 15, LNCS 4564, Online Communities and Social Computing, edited by Douglas Schuler
- Volume 16, LNAI 4565, Foundations of Augmented Cognition 3rd Edition, edited by Dylan D. Schmorrow and Leah M. Reeves
- Volume 17, LNCS 4566, Ergonomics and Health Aspects of Work with Computers, edited by Marvin J. Dainoff

I would like to thank the Program Chairs and the members of the Program Boards of all Thematic Areas, listed below, for their contribution to the highest scientific quality and the overall success of the HCI International 2007 Conference.

Ergonomics and Health Aspects of Work with Computers

Program Chair: Marvin J. Dainoff

Human Interface and the Management of Information

Program Chair: Michael J. Smith

Jiye Mao, P.R. China
Fiona Nah, USA
Shogo Nishida, Japan
Leszek Pacholski, Poland
Hidekazu Yoshikawa, Japan
Li Zheng, P.R. China
Bernhard Zimolong, Germany

Human-Computer Interaction

Program Chair: Julie A. Jacko

Sebastiano Bagnara, Italy
Jianming Dong, USA
John Eklund, Australia
Xiaowen Fang, USA
Sheue-Ling Hwang, Taiwan
Yong Gu Ji, Korea
Steven J. Landry, USA
Jonathan Lazar, USA
V. Kathlene Leonard, USA
Chang S. Nam, USA
Anthony F. Norcio, USA
Celestine A. Ntuen, USA
P.L. Patrick Rau, P.R. China
Andrew Sears, USA
Holly Vitense, USA
Wenli Zhu, P.R. China

Engineering Psychology and Cognitive Ergonomics

Program Chair: Don Harris

Kenneth R. Boff, USA
Guy Boy, France
Pietro Carlo Cacciabue, Italy
Judy Edworthy, UK
Erik Hollnagel, Sweden
Kenji Itoh, Japan
Peter G.A.M. Jorna, The Netherlands
Kenneth R. Laughery, USA
Nicolas Marmaras, Greece
David Morrison, Australia
Sundaram Narayanan, USA
Eduardo Salas, USA
Dirk Schaefer, France
Axel Schulte, Germany
Neville A. Stanton, UK
Andrew Thatcher, South Africa

Universal Access in Human-Computer Interaction

Program Chair: Constantine Stephanidis

Julio Abascal, Spain
Ray Adams, UK
Elizabeth Andre, Germany
Margherita Antona, Greece
Chieko Asakawa, Japan
Christian Bühler, Germany
Noelle Carbonell, France
Jerzy Charytonowicz, Poland
Pier Luigi Emiliani, Italy
Michael Fairhurst, UK
Zhengjie Liu, P.R. China
Klaus Miesenberger, Austria
John Mylopoulos, Canada
Michael Pieper, Germany
Angel Puerta, USA
Anthony Savidis, Greece
Andrew Sears, USA
Ben Shneiderman, USA
Christian Stary, Austria
Hirotada Ueda, Japan

Gerhard Fischer, USA
Jon Gunderson, USA
Andreas Holzinger, Austria
Arthur Karshmer, USA
Simeon Keates, USA
George Kouroupetroglou, Greece
Jonathan Lazar, USA
Seongil Lee, Korea
Jean Vanderdonckt, Belgium
Gregg Vanderheiden, USA
Gerhard Weber, Germany
Harald Weber, Germany
Toshiki Yamaoka, Japan
Mary Zajicek, UK
Panayiotis Zaphiris, UK

Virtual Reality

Program Chair: Randall Shumaker

Terry Allard, USA
Pat Banerjee, USA
Robert S. Kennedy, USA
Heidi Kroemker, Germany
Ben Lawson, USA
Ming Lin, USA
Bowen Loftin, USA
Holger Luczak, Germany
Annie Luciani, France
Gordon Mair, UK
Ulrich Neumann, USA
Albert "Skip" Rizzo, USA
Lawrence Rosenblum, USA
Dylan Schmorrow, USA
Kay Stanney, USA
Susumu Tachi, Japan
John Wilson, UK
Wei Zhang, P.R. China
Michael Zyda, USA

Usability and Internationalization

Program Chair: Nuray Aykin

Genevieve Bell, USA
Alan Chan, Hong Kong
Apala Lahiri Chavan, India
Jori Clarke, USA
Pierre-Henri Dejean, France
Susan Dray, USA
Paul Fu, USA
Emilie Gould, Canada
Sung H. Han, South Korea
Veikko Ikonen, Finland
Richard Ishida, UK
Esin Kiris, USA
Tobias Komischke, Germany
Masaaki Kurosu, Japan
James R. Lewis, USA
Rungtai Lin, Taiwan
Aaron Marcus, USA
Allen E. Milewski, USA
Patrick O'Sullivan, Ireland
Girish V. Prabhu, India
Kerstin Röse, Germany
Eunice Ratna Sari, Indonesia
Supriya Singh, Australia
Serengul Smith, UK
Denise Spacinsky, USA
Christian Sturm, Mexico
Adi B. Tedjasaputra, Singapore
Myung Hwan Yun, South Korea
Chen Zhao, P.R. China

Online Communities and Social Computing

Program Chair: Douglas Schuler

Chadia Abras, USA
Lecia Barker, USA
Amy Bruckman, USA
Peter van den Besselaar, The Netherlands
Peter Day, UK
Fiorella De Cindio, Italy
John Fung, P.R. China
Michael Gurstein, USA
Tom Horan, USA
Piet Kommers, The Netherlands
Jonathan Lazar, USA
Stefanie Lindstaedt, Austria
Diane Maloney-Krichmar, USA
Isaac Mao, P.R. China
Hideyuki Nakanishi, Japan
A. Ant Ozok, USA
Jennifer Preece, USA
Partha Pratim Sarker, Bangladesh
Gilson Schwartz, Brazil
Sergei Stafeev, Russia
F.F. Tusubira, Uganda
Cheng-Yen Wang, Taiwan

Augmented Cognition

Program Chair: Dylan D. Schmorrow

Kenneth Boff, USA
Joseph Cohn, USA
Blair Dickson, UK
Henry Girolamo, USA
Gerald Edelman, USA
Eric Horvitz, USA
Wilhelm Kincses, Germany
Amy Kruse, USA
Lee Kollmorgen, USA
Dennis McBride, USA
Jeffrey Morrison, USA
Denise Nicholson, USA
Dennis Proffitt, USA
Harry Shum, P.R. China
Kay Stanney, USA
Roy Stripling, USA
Michael Swetnam, USA
Robert Taylor, UK
John Wagner, USA

Digital Human Modeling

Program Chair: Vincent G. Duffy

Norm Badler, USA
Heiner Bubb, Germany
Don Chaffin, USA
Kathryn Cormican, Ireland
Andris Freivalds, USA
Ravindra Goonetilleke, Hong Kong
Anand Gramopadhye, USA
Sung H. Han, South Korea
Pheng Ann Heng, Hong Kong
Dewen Jin, P.R. China
Kang Li, USA
Zhizhong Li, P.R. China
Lizhuang Ma, P.R. China
Timo Maatta, Finland
J. Mark Porter, UK
Jim Potvin, Canada
Jean-Pierre Verriest, France
Zhaoqi Wang, P.R. China
Xiugan Yuan, P.R. China
Shao-Xiang Zhang, P.R. China
Xudong Zhang, USA

In addition to the members of the Program Boards above, I also wish to thank the following volunteer external reviewers: Kelly Hale, David Kobus, Amy Kruse, Cali Fidopiastis and Karl Van Orden from the USA, Mark Neerincx and Marc Grootjen from the Netherlands, Wilhelm Kincses from Germany, Ganesh Bhutkar and Mathura Prasad from India, Frederick Li from the UK, and Dimitris Grammenos, Angeliki Kastrinaki, Iosif Klironomos, Alexandros Mourouzis, and Stavroula Ntoa from Greece.

This conference could not have been possible without the continuous support and advise of the Conference Scientific Advisor, Prof. Gavriel Salvendy, as well as the dedicated work and outstanding efforts of the Communications Chair and Editor of HCI International News, Abbas Moallem, and of the members of the Organizational Board from P.R. China, Patrick Rau (Chair), Bo Chen, Xiaolan Fu, Zhibin Jiang, Congdong Li, Zhenjie Liu, Mowei Shen, Yuanchun Shi, Hui Su, Linyang Sun, Ming Po Tham, Ben Tsiang, Jian Wang, Guangyou Xu, Winnie Wanli Yang, Shuping Yi, Kan Zhang, and Wei Zho.

I would also like to thank for their contribution towards the organization of the HCI International 2007 Conference the members of the Human Computer Interaction Laboratory of ICS-FORTH, and in particular Margherita Antona, Maria Pitsoulaki, George Paparoulis, Maria Bouhli, Stavroula Ntoa and George Margetis.

Constantine Stephanidis
General Chair, HCI International 2007

Preface

This volume of the proceedings from HCII 2007 contains papers presented at the 7th International Conference on Engineering Psychology and Cognitive Ergonomics. The book contains contributions from approaching 250 authors from 18 countries. It is divided into four sections: Cognitive and Affective Issues in User Interface Design; Cognitive Workload and Human Performance; Cognitive Modeling and Measuring; and Safety Critical Applications and Systems. However, the contributions address a wider range of issues, and it has been a challenge to group them into Parts.

I would like to thank all the authors for their contributions to this book. It almost goes without saying that without their efforts, neither the conference nor this volume would have been possible. I would also like to extend my thanks to all the Program Board members who contributed their time and effort to the selection and reviewing of the papers. Finally, I must thank Constantine Stephanidis, Gavriel Salvendy and the HCII 2007 conference organization team as a whole. Their organization of the conference has been first rate and the effort that they have put in to produce these proceedings has been immeasurable. Without them, none of this would have been possible.

Don Harris, Editor

HCI International 2009

The 13th International Conference on Human-Computer Interaction, HCI International 2009, will be held jointly with the affiliated Conferences in San Diego, California, USA, in the Town and Country Resort & Convention Center, 19-24 July 2009. It will cover a broad spectrum of themes related to Human Computer Interaction, including theoretical issues, methods, tools, processes and case studies in HCI design, as well as novel interaction techniques, interfaces and applications. The proceedings will be published by Springer. For more information, please visit the Conference website: http://www.hcii2009.org/

General Chair
Professor Constantine Stephanidis
ICS-FORTH and University of Crete
Heraklion, Crete, Greece
Email: program@hcii2009.org

Table of Contents

Part II: Cognitive Workload and Human Performance

Part III: Cognitive Modeling and Measuring

Part IV: Safety Critical Applications and Systems

Part I

Cognitive and Affective Issues in User Interface Design

Designing Human Computer Interfaces for Command and Control Environments

Amardeep Aujla, Neville A. Stanton, Daniel P. Jenkins, Paul M. Salmon, Guy H. Walker, and Mark S. Young

School of Engineering and Design, Brunel University, Uxbridge Middlesex, UB8 3PH
daniel.jenkins@brunel.ac.uk

Abstract. This paper will introduce the human factors command and control test bed developed at Brunel University. The system was developed to facilitate experiments into command and control within a military context. The purpose of the system is to support experimentation, it is not intended to represent a product that could be used in the field by the MoD. The test bed developed represents a controlled environment allowing the manipulation of individual variables. The manipulation of these variables allows researchers to address fundamental human factors questions emerging from the transition form an analogue paper based planning process to a digital network enabled process. Areas of particular interest for this system include collaborative working, distributed command centres, the flow of information as well as changes to the command hierarchy. The system consists of a number of commercial off the shelf products synthesised by a bespoke planning application.

Keywords: Command and Control; Test Bed; Experimental Environment; Software Development.

1 Introduction

The system described in this paper has been developed for the Human Factors Integration Defence Technology Centre (www.HFIDTC.com) at Brunel University. The system's raison d'être is to facilitate the manipulation of variables in a controlled repeatable environment. The system is designed to support other more naturalistic research currently underway within the consortium to address fundamental questions on the way forward for the transition to a Network Enabled Capability (NEC) system for the British military.

Command and Control or C2 is defined by Builder, Bankes and Nordin (1999) as:

"The exercise of authority and direction by a properly designated [individual] over assigned [resources] in the accomplishment of a [common goal]. Command and control functions are performed through an arrangement of personnel, equipment, communications, facilities, and procedures which are employed by a [designated individual] in planning, directing, coordinating, and controlling [resources] in the accomplishment of the [common goal]." (p. 11).

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 3–9, 2007.

In other words command and control is the product of combining command (authority) with control (the means to assert this authority), these are the emergent properties of, “unity of effort in the accomplishment of a [common goal]” (Jones, 1993) and, “decision superiority” (DoD, 1999p. 28).

Beyond the descriptive level, command and control is a collection of functional parts that together form a functioning whole. Command and control is a mixture of people and technology, typically dispersed geographically. It is a purposeful intelligently adaptive endeavour representing progress towards a defined outcome. Intelligent adaptiveness requires responses to externally generated input events within a finite and specified period (Young, 1982). In possessing these attributes, command and control can be characterised with reference to, and understood from the following modelling perspectives, as:

- An (open) system of interacting parts,
- A socio technical system of human and non-human agents and artefacts,
- A distributed system,
- A real time system, and
- An intelligent system.

The system discussed in this paper has been developed based on the collective learning gained from extensive observations of command and control systems in the fields of Army, Navy, Air force, Police Force, Fire service, Air traffic Control, National Grid and National Rail. (Houghton et al (2006); Stanton & Baber (2006); Stanton et al (2007); Walker et al (2006))

The system allows investigation into the effects on a team’s command structure and its ability to achieve its goals. At present the system represents a chain of command at three levels (Gold, Silver and Bronze). The terms Gold, Silver and Bronze have been taken from the emergency services to represent three levels of command. These terms were selected over military specific terms in order to allow research to be conducted at a number of levels of the command structure. The role of each of the command levels can be significantly changed by adding and removing displayed information as well as the level of functionality. This ability of the system allows experimentation into the optimal display requirements for each level for any given situation.

Fig. 1. Hierarchical Command structureshows an example of a traditional hierarchal structure here information is passed up from bronze to silver then aggregated and send to gold. Information is also sent down from gold to silver then disseminated to bronze. With the new capabilities of NEC it is now possible to send information in almost any conceivable way as shown in **Fig. 2.** Hierarchical Command structure, information can now easily be copied and distributed directly to gold, or via a peer to peer network.

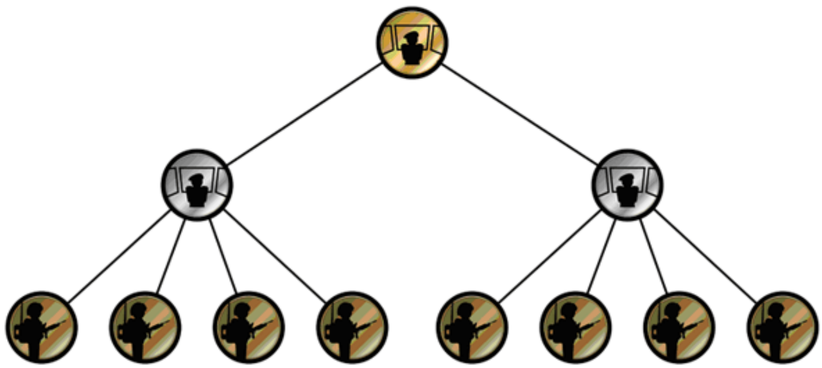


Fig. 1. Hierarchical Command structure

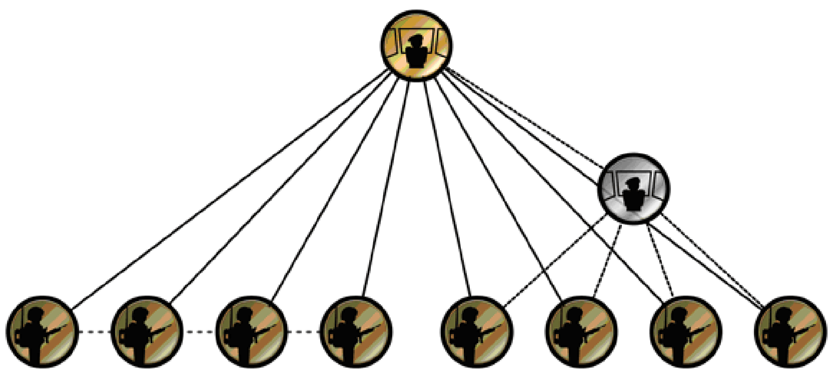


Fig. 2. Hierarchical Command structure

2 The System

The system has been developed to be fully mobile, it takes advantage of existing 3G networks allowing experiments to be conducted globally anywhere a mobile phone network can be found. A number of experiments have been conducted at Brunel University's Uxbrigdge campus, this environment is a controllable setting it provides a realistic and, therefore, ecologically valid urban battlespace landscape. The campus covers an approximately rectangular area of 50 hectares, with an elevation of 7.5 metres and no significant gradient. The campus is laid out with 20 definable structures (mainly concrete) ranging in height from approximately 3 metres to 20 metres (1 story to 8 stories respectively). The land adjacent to and between the structures is covered with hard paving and grass. The total battlespace is bounded by a perimeter road on all boundary faces, beyond which is chain link fencing on the South and North boundaries, a public road on the West boundary and a small river on the East boundary. A 3D representation of the total battlespace is provided in **Fig. 3.** 3D model of campus environment Technically the test bed is made up of a number of

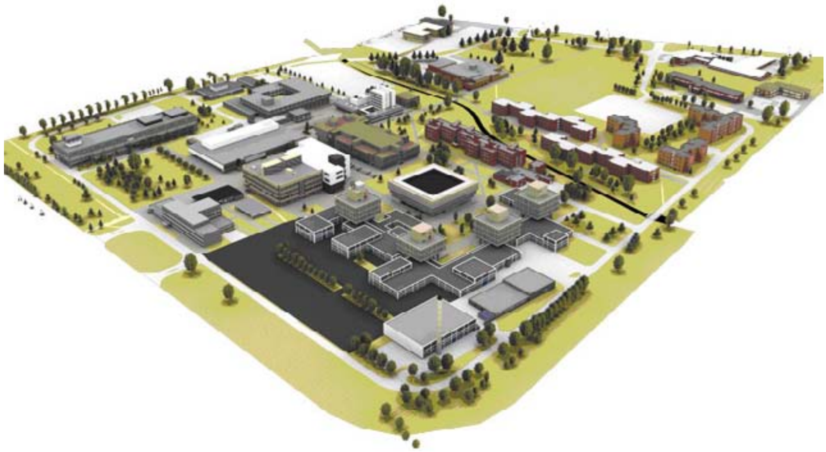


Fig. 3. 3D model of campus environment

Fig. 4. Layout of Gold Command Centre

facilities the gold command centre contains three large (3m by 2.25m) Stewart screens with rear projection for the peripheral screens and a front projector for the centre screen. Six Eiki LC-SX4Li projectors are projected onto the screens, the periphery screens are rotated by 45 degrees from the centre screen creating an immersive environment. The centre projector is 4.27m from the ground whilst the

rear projectors are 1.30m. This allows the user to get very close to the screen without obscuring the projected image (see **Fig. 4.** Layout of Gold Command Centre). Green et al (2005) offer a more detailed description of the system and its development.

The information displayed on the screen is completely reconfigurable typically information will be supplied by a number of field agents (bronze units) and represented on the map. Typical information will include their position as well as annotations for the agents and the area around them.

Silver units are presented with completely reconfigurable displays; the information is presented to the unit on a standard laptop equipped with a 3G data card. Bronze units are equipped with handheld PDAs (O2 XDA Execs) running Windows Mobile 5.

Fig. 5. Brunel Gold Command Centre

In order to expedite the development time of the command software running at the three levels it was decided to investigate a number of existing "commercial off the shelf" (COTS) software packages. After some preliminary investigation it was decided that the most promising option to investigate further was Google Earth.

Google Earth is a free downloadable piece of software that has a number of benefits for tracking the position of objects in a 3D streamed data version of the real world. After a very short investigation it was clear that the software supported many of the required functions. Like many other internet based programs Google Earth contains satellite imagery of varying quality for the entire globe. Google Earth however contains some clever algorithms that link this data together and stream it to

the user. It is possible to quickly navigate around by a number of methods, viewing the scene at a number of zoom levels. Additional information and control can be added by getting the software to read and write very simple text files called .KML. These are in many ways very similar to HTML. The files can be easily added to a scene by dragging and dropping or by an automated link allowing information to be shared between concurrent users.

In order to add in further functionality such as support for the combat estimate planning process an additional program is required to run along side Google Earth. This program needs to be able to send and receive as much information as possible to allow communication between the programs to be complete and seamless.

The first stage of this process is to develop a piece of software as a proof of concept that demonstrates communication between the two programs. **Fig. 5.** Brunel Gold Command Centre shows the two programs working with Google Earth on the left hand side and the new bespoke software on the right hand side. The icons can be seen to be positioned in the same location on both maps. When moving from a 2D static paper based system to a 3D dynamic system there are a lot of considerations to be made. The C4I environment is very different to the traditional 2D birds eye view maps used that are marked up with symbols printed on transparencies stored in large folders. The 3D environment by its very nature can be viewed from any angle and at various zoom distances. This brings with it many advantages, the commander can view the battlefield in a traditional birds eye view or can view the battle-space from a soldiers eye view. The map modelled at Brunel is of a 3D texture mapped environment that is dynamic, the map is automatically updated with time as units move and other information is received. The new technology brings with it lots of additional capabilities that need to be carefully considered before they are implemented. The ability to model future and past activities as animations now exists as well as true real time updating of information, careful consideration has to be placed on how this information is represented and who should receive it. Around the world there have been a number of developments of military symbology for computer based command and control applications. (Albinsson & Fransson 2002; Object Raku Technology 2005; Mouat 2005). The display of warfighting symbology has evolved from a static, manual operation to include fully automated computer generation. A representation method is required that can be tranmitted over a number of communication channels at low bandwidth and rapidly regenerated and understood at the other end. “The standardisation of warfighting symbology plays an integral role in achieving interoperability during joint service operations. While the primary focus of this standardisation is the electronic generation of symbology, this effort must also support those mission requirements where symbology is hand drawn by the warfighter.” (US DOD 1999). The fact that the domain modelled is dynamic lends itself very well to a C4I representation. Albinsson (2005) comments that the system as a whole will, to a great extent, change states without a particular user’s intervention. Here for the first time the processof updating the map has become automated, the action of automating this processs brings with it many additional considerations, for the first time the human is taken out of the loop and the ability to apply a ‘sanity check’ is removed.

3 Conclusions

This document has outlined the development of the Brunel Command and Control System.

The requirements have been born out of documents created by observation and analysis of multiple military, emergency services and civilian command and control scenarios. The system has and will continue to be developed in an iterative fashion incorporating new currently unanticipated levels of functionality as they are identified by future analysis and experimentation.

The paper has introduced a number of areas to be investigated further and raises a number of issues that need to be addressed in the move from a traditional static paper based representation of the domain to a dynamic digital representation. Work is currently ongoing on the development of the system in line with a number of experimental studied.

Acknowledgement. This research from the Human Factors Integration Defence Technology Centre was part-funded by the Human Sciences Domain of the UK Ministry of Defence Scientific Research Program. Any views expressed are those of the authors and do not necessarily represent those of MOD or any other UK government department**.**

References

1. Albinsson, P.-A., Fransson, J., Representing Military Units Using Nested Convex Hulls – Coping with Complexity in Command and Control. In: The 1st Swedish-American Workshop on Modeling and Simulation. Orlanda, USA (2002)
2. Builder, C.H., Bankes, S.C., Nordin, R.: Command Concepts: A Theory Derived from the Practice of Command and Control. Rand, Santa Monica, CA (1999)
3. DOD: Department of defence interface standard, common warfighting symbology MIL-STD-2525B (1999) http://symbology.disa.mil/symbol/mil-std.html
4. Green, D., Stanton, N.A., Walker, G.H., Salmon, P.: Using wireless technology to develop a virtual reality command and control centre. Virtual Reality 8, 147–155 (2005)
5. Houghton, R.J., Baber, C., McMaster, R., Stanton, N.A., Salmon, P., Stewart, R., Walker, G.: Command and control in emergency services operations: A social network analysis. Ergonomics 49(12-13), 1204–1225 (2006)
6. Wickens, C.D., Merwin, D.H., Lin, E.L.: Implications of Graphics Enhancements for the Visualisation of Scientific Data; Human Factors; 36(1) 44-61 (1994) Object Raku Technology (2005) War symbology digital version of MIL-STD-2525B http://www.objectraku.com/war_symbology/04war_symbology.html
7. Stanton, N.A., Baber, C.: The Ergonomics of command and control. Ergonomics 49(12-13), 1131–1138 (2006)
8. Stanton, N.A., Baber, C.: The Ergonomics of command and control. Ergonomics 49(12-13), 1131–1138 (2006)
9. Walker, G.H., Stanton, N.A., Gibson, H., Baber, C., Young, M., Green. D.: Analysing the role of communications technology in C4i scenarios: A distributed cognition approach. Journal of Intelligent Systems 15(1-4), 299–328 (2006)
10. Young, S.J.: Real time languages: Design and Development. Ellis Horwood, Chichester (1982)

Perceived Complexity and Cognitive Stability in Human-Centered Design

Guy A. Boy

European Institute of Cognitive Sciences and Engineering
EURISCO International, 4 avenue Edouard Belin, 31400 Toulouse, France
guy.boy@eurisco.org

Abstract. Perceived complexity is analyzed in conjunction with cognitive stability in the context of potential use in human-centered design. These human factors are useful in the process of understanding co-adaptation of people and technology, and consequently for the evaluation of the maturity of a product. Expertise and experience play an important role in the definition and refinement of these two concepts. This paper presents a first account of such concepts in the context of aircraft cockpit design.

Keywords: perceived complexity, cognitive stability, maturity, expertise, experience, co-adaptation, esthetics, natural versus artificial, knowledge, know-how, skills, redundancy, cognitive support, safety, performance, comfort.

1 Introduction

This paper presents an analysis of cognitive stability and perceived complexity in human-machine interaction, where the machine is software-intensive. Both concepts are complementary. Perceived complexity is more related to the "gulf of evaluation", and cognitive stability to the "gulf of execution", in Norman's terminology [19] [20]. Cognitive stability is related to various principles such as simplicity (and its counterpart complexity), observability and controllability, and redundancy. Even if adaptation is an asset of human beings, their life is better when technology is adapted to them. In fact, the real issue in human-centered design is to better understand *co-adaptation* of people and technology [4] in the perspective of increasing cognitive stability. Cognitive stability is defined taking the physical metaphor of passive and active stability that respectively involves static and dynamic complexity.

Our technological world never stops producing new technology that induces new practices, therefore people attempt to adapt in order to reach a reasonable level of cognitive stability. We observe that people may, or may not, have difficulty to learn and maintain these practices. Such difficulty is directly related to system complexity and emerging cognitive instability. The main problem arises when we try to understand what system complexity really is. System complexity will be seen from the perspective of a user who has to adapt to a new system or technology. Consequently, the appropriate concept that will be studied is *perceived complexity*. A system may be internally complex but very well designed, and provide a very easy

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 10–21, 2007.

user experience. Conversely, a technology that is not mature enough may cause an overload of adaptation that requires extensive training on this technology. Therefore, *maturity* of both technology and practices is a concept that deserves to be further investigated. In the context of life-critical systems, certification of new technology is required [8]. Certification criteria are built around two main kinds of principles, design principles that are concerned with system stability and use principles that are concerned with cognitive stability.

"*Errare humanum est*" (i.e., making errors is human) is a very old maxim that is still very much verified and used. Its corollary "*Perseverare diabolicum*" (i.e., persisting is diabolical) should also be remembered. For the last three decades, we have studied human errors in the cockpit of commercial aircraft. Well-informed statistics, such as the ASRS [2], report that about 75% of incidents are caused by pilots' *human errors*. Several authors proposed various kinds of human error taxonomies [1] [12] [15] [18] [22]. Hollnagel talks about erroneous actions that can be initiated or terminated too early, too late under incorrect conditions, for incorrect reasons, or not at all [13]. The priority assigned to an action can also be a factor of human error.

But how can a designer or engineer take into account potential human errors at design time? A designer is a creator or an architect of a novel artifact. An engineer is a builder. It happens that some people have both of these jobs. Let the term "design" denote both of them. A designer then needs to understand human reliability in order to prevent surprises. Even if design is an incremental trial-and-error process, the best early design is likely to lead to the best product in the end. This is why there is a need for an underlying theory of both cognitive stability and perceived complexity.

The first part of this paper provides an account of the concept of cognitive stability using the physical stability background and the latest development in human reliability and adaptation. The second part is devoted to the analysis of perceived complexity. The third part shows how human-centered design can be improved by taking into account cognitive stability and perceived complexity. In the balance of the paper, a discussion is started on related issues.

2 Cognitive Stability

2.1 Passive and Active Stability

Stability is a concept that physicists know well. The pendulum is a good example of passive stability. A pendulum consists of a mass attached by a string to a pivot point. When the mass is moved aside, it tends to return to its initial state. We talk about stable balance or equilibrium. The same kind of phenomenon happens when you through a ball in a bowl, it stabilizes at the bottom of the bowl… and stays in the bowl. However, if you turn the bowl and put the ball on the top of its convex side, then the ball falls and is no longer on the bowl. We talk about instable balance or equilibrium. In the same way, plate spinning consists in maintaining a plate on a pole without it falling off. Plate spinning relies on the gyroscopic effect that you need to permanently induce from the basis of the pole. Circus acrobats are able to maintain several spinning plates on the top of several poles at the same time. This is active stability.

A chair requires three legs that are not aligned to be passively stable. However, why are human beings able to stay erect on two legs? This is because we constantly compensate balance errors. We stay erect because we are active, often unconsciously! We learned this kind of active stability in our very early childhood. Taking this metaphor seriously, an erroneous action deviates from a stable state either to return to it or diverge from it. The former defines cognitive stability and the latter cognitive instability.

There are human-machine systems that provide cognitive stability, at least in a specific context of use. For example, there are bicycles that are physically stable from the intrinsic design of their forks, i.e., when you move the handlebars sideways (but not too much) they tend to re-stabilize on the trajectory. Others are not. In the same way, most current spelling checkers tend to automatically correct typos, or at least suggest appropriate corrections, without disturbing the generation of a text. They tend to make text processing more comfortable and efficient (i.e., the text is [almost] free of spelling mistakes). This kind of passive stability is related to the concept of error tolerance, i.e., users may make errors that are tolerated, in the sense that they can be either automatically corrected or users have the awareness and means to recover from them.

Conversely, there are systems that induce cognitive instability. For example, the carpenter without harness or appropriate equipment who is walking on a beam located several meters above the ground takes a risk. If he makes an error, is unbalanced and falls, there is no recovery means. Professional carpenters learn to actively stabilize their walking on beams. In addition, they may wear appropriate equipment such as harness. In the same way, current computers have many layers of software that users cannot and don't want to understand. Sometimes, when you push a button, the computer screen provides a message saying "Processing…", without anything else. As a result, impatient people start to believe that the computer is down. They usually need to act, and press other keys or buttons to check if the computer is still active. What they do at this point is filling a buffer of commands that will be active when the "Processing…" is finished! They then observe unexpected behavior of the machine, and they actually may continue to press keys and buttons, and so on. This is a diverging process induced by a bad design. This is induced cognitive instability that is directly related to the concept of error resistance. Usability engineering has already provided solutions for this kind of example [17].

You can see that cognitive instability is directly linked to expertise and experience either from designers or users. Experience strongly contributes to decrease perceived complexity. Of course, we would like to design cognitively stable machines for experienced people as well as for novices. Both users and designers should make a permanent co-adaptation in order to reach an acceptable stability of the overall human-machine system.

2.2 Intentional Versus Reactive Behavior

Someone may be in an instable situation intentionally or not. For example, a funambulist walks or rides on a rope intentionally. He or she learns how to keep a reasonable balance, i.e., how to compensate for his or her unbalance errors. He or she is an expert. Novices start with long poles in order to create an artificial inertia that

almost automatically maintains an acceptable balance. As training advances, the size of the pole may decrease. This physical prosthesis is incrementally replaced by a cognitive skill through learning and experience. More generally, skills and knowledge help compensate physical instability, as cognitive instability. Of course, such skills and knowledge could be transferred to a machine in the form of automation. Activities such as funambulism are goal-driven in a very well known environment.

Conversely, a population experiencing an earthquake or a tsunami has to react to them if they can. Reacting in such cases consists in protecting oneself. People's activities in such cases are event-driven in an unknown environment. They are untrained. Since people do not usually have any skills to react, they either use high-level cognition (Rasmussen's knowledge-based behavioral level) or react in a random way. They may also rely on someone who knows and whom they trust.

It is useful that designers keep in mind this distinction between intentional and reactive behavior, as well as the distinction between known and unknown environment, when they are involved in a design task. Users may be either experts or novices with respect to normal or abnormal situations. In some abnormal or emergency situations, even experts and experienced people may return to a basic novice behavior. In other words, design should adapt for both goal-driven and event-driven activities. I fully admit that these things are not necessarily obvious in the early stages of the design process. However, they may guide to set up experimental tests and interpret their results in order to incrementally improve the initial design.

2.3 Human Adaptation and Expertise

As already said, cognitive stability is a matter of expertise when the machine only offers unstable possibilities. Flying an airplane requires an expertise that consists in using skills coming from a long and intense training and experience. On modern commercial aircraft, automation liberates pilots from a lot of compensations that their predecessors used to perform. However, as stated above "*errare humanum est*" is still true, and they make errors using the resulting automated machine. A new adaptation has always to be done, not on the same issues as before, but on new ones. Automation has incrementally introduced novel issues such as situation awareness for example. Why? This is just because the layers of software, introduced between the users and the "initial" mechanical machine, contributed to remove cues that users were using to control the machine before. These software agents take care of these cues, sometimes without reporting to users. Designers then need to understand what new cues users need. Users have become managers of software agents in the same way someone who gets a promotion is now responsible for a team of agents working for him or her.

Cognitive stability then takes another face. Users as managers of machine agents state the problem of human-machine communication, cooperation and coordination. We were before in a context of manipulation of tools, we are now in the context of collaboration with agents. This is a fundamentally different situation. Adaptation is no longer sensor-motoric, but rather cognitive. Cognitive stability therefore requires a better understanding of how agents communicate among each other. Three models of agent communication were proposed in previous papers [5] [14]: supervision, mediation and cooperation by mutual understanding. Cognitive stability then becomes socio-cognitive stability.

3 Perceived Complexity

3.1 Connectivity, Separability and Perceived Complexity

We saw that cognitive stability is related to expertise when a machine is badly designed, but it might also be the case when this machine is internally complex and not mature enough. This was the case of cars in the early times when a driver was also a mechanic. This was the case of computer users until computing power enabled the construction of easy-to-use graphical interfaces. Using tapes and cards to program computers was not an easy thing to do, only experts were able to perform this kind of practice. It took visionaries, such as Douglas Engelbart who invented the mouse and contributed in the first concrete design of hypertext and internet, to propose and test new devices and architectures to develop what we know today.

A system is made of parts that are interlinked with each other, and when these links cannot be broken without destructing the "life" of the whole, then the system will be said to be complex. Conversely, when two parts can be dissociated and processed independently without harming the whole, the system will be said to be separable. Therefore, we see that system complexity is strongly related to *connectivity*. A living system such a human being cannot be easily broken into parts. Parts are interconnected with several dependencies. Some of these dependencies are vital, and others are less vital. Living systems are usually stable because of some of their redundant parts. The respiratory system is redundant for example, i.e., people are able to live with only one lung. There are obvious consequences on the level of effort that people with only one lung are able to perform, but the vital oxygenation of the blood is still working fine.

Attempting to understand system complexity is then a matter of finding out the salient parts and their interrelations. Designers and users of a system may not see the same parts and interrelations, just because they do not have the same tasks to perform with respect to the system. They do not "separate" the system in the same way because they do not have to understand the logic of the system in the same way. Complexity is intimately related to *separability*. When a doctor administrates a medication to a patient, he or she has to know the secondary effects of this medication, i.e., acting on a part may have an effect on other parts. When a part, e.g., the respiratory system, is failing, medication is usually provided to treat the disease, but this medication may have an impact on other parts of the body, i.e., the whole system. Of course, we will always attempt to separate what is separable in order to simplify! But there will be an end to this separability process. There are "atomic" parts that are not at all separable. These atomic parts live by themselves as a whole. The problem is then to figure out how complex they are when we handle them. People facing "inseparable" complexity do not usually react in a logico-mathematical way but use heuristics. This inseparable complexity mainly results from the use of poorly integrated automation and multiple unarticulated layers of software. Sometimes, people may start thinking something is too complex without even trying and end-up not being able to realize a task simply by the fact that they persuade themselves that they can't do it. This is the Acquired Incapacity Syndrome (AIS).

Human-machine complexity is a matter of *perceived complexity* that cannot be described by an axiomatic approach except when the system can be separated into

parts that can be worked separately. Complexity perceived by the user was previously described in comparison with internal complexity of the machine [11]. We need to make a distinction between perceived complexity and difficulty, e.g., some tasks may remain difficult even when the user has acquired experience and has been trained.

3.2 Redundancy

Signs that appear to be superfluous in the understanding of a message are said to be redundant. The following sentence is plural-redundant (three times): “Dolphins are good swimmers”. If by mistake you drop an “s” then the reader will still understand that the sentence is plural. By repeating a word, we increase the effect of this word such as: “This is very very good!” If you delete one of the “very” then the sentence will still emphasize the same kind of idea. Similarly, the space shuttle has five computers, three inertial systems and two independent systems that compute its position according to the stars. If one of the computer/system fails the other(s) will still manage to keep the shuttle working. Human beings have their own redundancy in order to survive. More commonly, when you use a computer, you need to be redundant – you need to make backups of your files in order to work safely.

People are equipped with several redundancies to insure their stability in their environment. For example, peripheral vision is a permanent redundancy to central vision. Peripheral vision is useful for spatial orientation and motion cues, while central vision is useful for detailed imagery and color perception. Central vision provides the “what” of the scanned target, while peripheral vision provides the “where”. The “where” can be said to be redundant to the “what”. We have run a series of experiments during the late 90s with aircraft pilots to better understand how they use procedures and checklists, and why in some cases they are do not use them [10]. One of the major results was that 65% of the pilots on 245 involved in the experiment, did not use a procedure item when they did not know why they had to, i.e., they where provided with the “what” and not the “why”. In this case, redundancy means complementarity.

More generally, the supervision of highly automated systems requires redundant information on the “why the system is doing what it does”, “how to obtain a system state with respect to an action using control devices”, “with what other display or device the current input/output should be associated” and “when it will provide the expected information” in order to increase insight, confidence, and reliability. The paradox is that by increasing redundancy, and therefore system complexity, perceived complexity tends to decrease.

3.3 Cognitive Support

Error tolerance and error resistance systems are usually useful redundancy. Redundancy is cognitive support. Error tolerance is always associated to error recovery. There are errors that are good to make because they foster awareness, recovery and learning. However, recovery is often difficult, and sometimes impossible, when appropriate resources are not available. These appropriate resources are cognitive support to users. Whenever an action is reversible, the user can backtrack from an erroneous action, cognitive support should be available to correct

this error. For example, the UNDO function available on most software applications today provides a redundancy to users who detect typos and decide to correct them. Thus, making typos is tolerated, and a recovery resource is available. As we already saw, error resistance is, or should be, associated to risk. Error-resistance resources are useful in safety-critical systems when high risks are possible. They may not be appropriate in low-risk environments because they usually disturb task execution. In that case, they are bad cognitive support to users. For example, text processors that provide permanent automatic semantic grammar checking may disturb the main task of generating ideas.

People have cognitive functions to anticipate action, interact and recover from errors. These cognitive functions can be enhanced by the use of various categories of cognitive support. These categories may be related to the physical tool being used, the capacities of its user, the task being performed, and their organizational environment. The concurrent use of color, shape and grouping of related states is a classical example of cognitive support. People perceive both contrasts as well as similarities. Improbable information should be highlighted in order to anticipate possible surprises. Redundant feedback to user's actions is likely to improve interaction. Additional assistance to action taking should improve recovery.

People acting usually look for cognitive support consciously or unconsciously, in the same way they look for physical support, e.g., we may grab a ramp to climb stairs. When such support is not available, people build it either as good practice or as external devices supporting their activities. The development of situation patterns and processing habits such as systematic crosschecking is likely to improve anticipation, interaction or recovery. Training for error recovery is likely to improve cognitive stability.

A way to increase cognitive support is to automate the user interface by introducing some procedural action items into it. In domains where interaction is required in real-time, we know by experience that the number of commands, displays and entry fields should be reasonably small at any given point in time. These interaction devices and instruments just may not be the same at all times. This remark is crucial, i.e., "Context is important!" You don't need the same interface in two different contexts. This breaks the consistency rule that is commonly used in static environments such as text processing, i.e., "Be consistent lexically, syntactically, semantically!" However, people are pragmatic and they are "consistent pragmatically" also, i.e., they correlate objects to situations. Procedural interfaces [6] provide appropriate means to enhance this kind of correlation. They are explicitly based on the old ideas of the "Art of memory" [27]. Again, perceived complexity is likely to increase even if the user interface complexity increases.

4 Cognitive Stability and Perceived Complexity in Design

4.1 Incremental Nature of Design

A product in its environment cannot provide cognitive stability and perceived complexity without an appropriate method used during its design and development. Design is incremental by nature. Human-centered design involves the gathering of

various stakeholders including designers, human factors specialists, and end-users. End-users are not designers, but they should be part of the design process by being involved in formative evaluations.

The design of a new system, or more generally a technology, should start with a good idea! This idea has to be processed by an expert and experienced team who knows the application domain. This is why cognitive engineers should be closely involved in an application domain, or work closely with domain experts. In addition to expertise and experience, it must be acknowledge that to get this good idea realized into a product, it may take a fair amount of time. Why? Because refinement of a first good design is the most time-consuming part of product development. Refinement includes a large number of formative evaluations with end-users, and discovering what kind of new roles or jobs novel technology involves. It takes patience and continuous efforts!

The design process should then be guided by a modeling/documentation support in order to incrementally rationalize it. Incremental rationalization tends to improve the refinement process. This paper claims that, in human-centered design and development, human modeling is crucial to rationalize cognitive stability and perceived complexity.

4.2 The Importance of Human Modeling in Design

A great amount of work has already been done in human modeling for complexity analysis in *human-machine interaction* (HMI). This work started in *human engineering* (HE), *human-computer interaction* (HCI) and *cognitive engineering* (CE). HMI concerns factors induced when a human operator executes a task using a machine (or a tool). Related studies are on the analysis and assessment of behavioral and cognitive human factors. We make a distinction here between *human-machine interaction through computers* and HCI. HCI was initially developed within the context of office automation since the mid-1980s[1]. A large community emerged and is now expanding to other application domains such as car and aircraft automation. At the same time, other research and engineering communities investigated the integration of computers in various human-machine interaction situations. In the aviation domain, human factors specialists were interested in safety issues related to automation, and in particular in human errors and human reliability. Today, these various research and engineering communities tend to merge toward the development of unifying approaches.

Machines are becoming more and more computerized, i.e., computers are interface devices between the (mechanical) machine and the operator. The computer is a new tool mediating human-machine interactions. Such a mediating tool is called a *deep interface*. The physical interface is only the surface of this deeper interface and what global and local aspects they perceive. A current research topic is to better understand what operators need to know of this deeper interface. Should they only know the behavior of the physical interface? Should they understand most of the internal mechanisms of the deeper interface? How should the deeper interface represent and

---

[1] ACM-SIGCHI, for example, is one of the most well known community of research and practice in human-computer interaction.

transfer the behavior and mechanisms of the (mechanical) machine? Answers to these questions should guide designers to re-focus both the design of operations modes, and training.

Current work mainly consists in managing highly computerized systems, which leads to supervisory control, delegation, cooperation and coordination of artificial agents activities. Consequently, a new discipline has emerged, called *cognitive engineering* [7]. Cognitive models have been developed that take into account the new evolution of *human-machine interaction through computers*. Cognitive engineering is evolving at the same time as other sister-disciplines such as control theories, artificial intelligence, cognitive psychology, anthropology and sociology.

From a philosophical viewpoint, the issue of human-machine systems can be seen as whether the coupling is between the process and the computer or between the computer and the user. Current fly-by-wire aircraft, designed in the mid-eighties, are instances of this human-machine system model. The distinction between *amplification* and *interpretation* is important because it entails two completely different views of the role of the human in human-machine systems, hence also on the design principles that are used to develop new systems. In the amplification approach, the computer is seen as a tool. In the interpretation approach, the computer can be seen as a set of illusions re-creating relevant process or system functionalities.

The OZ pilot interface for example uses a deep conceptual model based on vision science, cognition and human-centered computing [23]. The OZ display removes the burden of scanning among flight instruments. It takes into account the fact that the human visual system is divided into two channels: the focal channel (central vision) and the ambient channel (peripheral vision). Conventional cockpits require tedious scanning mobilizing the focal channel, while the ambient channel is purely unused. OZ shows luminance discontinuities, i.e., moving lines and dots, that are resilient to one and/or two-dimensional optical and neurological demodulations. The resilience of conceptual primitives to demodulation allows them to pass information through both focal and ambient channels' optical and neurological filters [24] [25] [26]. These dynamic visual objects seem to improve cognitive stability and tremendously decrease perceived complexity by disambiguating some complex perceptual cues such as relative movements. One of the reasons is that OZ exploits continuous movements of objects meaningful to the pilot and, at the same time, makes clear distinctions among these objects. OZ affords pilots to use their complementary visual channels, and therefore promotes the cue of sensory redundancies.

4.3 Categorization of Cognitive Stability and Perceived Complexity Attributes

"Complexity refers to the internal workings of the system, difficulty to the face provided to the user -- the factors that affect ease of use. The history of technology demonstrates that the way to make simpler, less difficult usage often requires more sophisticated, more intelligent, and more complex insides. Do we need intelligent interfaces? I don't think so: The intelligence should be inside, internal to the system. The interface is the visible part of the system, where people need stability, predictability and a coherent system image that they can understand and thereby learn." [21]. This citation from Norman is very important today when we have layers and layers of software piled on top of each other, sometimes designed and developed

to correct previous flaws of lower layers. We commonly talk about patches. This transient way of developing artifacts does not show obvious maturity. The maturity of an artifact can be defined by its robustness, resilience and availability. As already said, it is always crucial to start with good high-level requirements that of course can be refined along the way. The problem comes when they are weak!

We have recently run a cycle of interviews and brainstorming with airline pilots on the way they see perceived complexity and cognitive stability. We have deduced the following series of attributes. What came first is the concept of mandatory training when the use of a system to be controlled is perceived as complex. It came out that adaptation is a necessary attribute of cognitive stability. The second concept concerned the affordance of displayed information, i.e., get the right information at the right time in the right format. This is also adaptation, but adaptation of the system this time, i.e., the design should take care of proper color contrast, adequate colors, and proper information and symbology. In addition, pilots insisted on the fact that crosschecking is key, especially using fly-by-wire aircraft, where one pilot needs to know what the other one is doing. Cognitive stability is based on both an emerging practice (know-how and procedures) and appropriate displays and controls. Dealing with the unpredicted in real-time is a matter of disturbance of cognitive stability, in particular pilots said that perceived complexity increases when they have to manage automation under time pressure, and in case of failure, it is almost always “hurry up”. Solutions often involve the execution of a minimal number of actions (vital actions) in minimal time, and possibility to manage action priorities (anticipation, preparation). In order to increase cognitive stability, such solutions should be either prepared in advance using procedural support or integrated in a procedural interface [6]. The same works for alarm management where cognitive support can be provided through either visual or auditory channels to facilitate anticipation, and therefore improve cognitive stability. As far as colors are concerned, the number of color codes cannot be larger that 7±2 (Miller’s law, [16]) and contrasts must be as strong as possible. Such factors contribute to decrease perceived complexity because they facilitate user’s working memory management.

Conducted interviews and brainstorming with experienced pilots (both test pilots and airline pilots) enabled the elicitation of the following perceived-complexity factors: expertise, visibility and affordances, social-cognition, uncertainty, alarm management, levels of automation culture, degree of explanation of system internal complexity, the level of operational assistance, the appropriateness of interaction cognitive functions involved, error tolerance, clarity and understandability of the language being used, flexibility of use, display content management, risk of confusion, assistance in high-workload situations, rapid re-planning, trust, technology and practice maturity, user-friendliness, ease of forgetting what to do, lack of knowledge, experience and training, usability, mode management, the another-function syndrome, multi-agent management, system feedback, the PC screen do-it-all syndrome, interruptions, information-limited attractors, conflicting information or diverging information, abnormal situations, extrinsic consistency, lack of flexibility, lack of design rationale availability, predictability, redundancy, information modality, information saturation, and situation awareness. This preliminary list has been extended and categorized into five classes with respect the AUTOS pyramid, i.e., Artifact, User, Task, Organization and Situation [3] [9].

5 Conclusion and Perspectives

Perceived complexity and cognitive stability are complementary concepts that are useful to articulate during the whole life cycle of an artifact (system or technology). The earliest during the design process is the best of course! However, these two concepts are also useful for later evaluation purposes. Cognitive stability has been described using the metaphor of stability in physics. Perceived complexity is incrementally defined from user experience by progressive refinements.

I strongly believe that there is no general human model that can help to solve all design problems to reach a satisfactory mature product. Concepts such as perceived complexity and cognitive stability are very useful to figure out what to measure and what to evaluate, and potentially guide design. They need to be further elaborated on concrete use cases in order to derive potential human factors measures.

Acknowledgements. Sébastien Giuliano greatly helped in the early development of the perceived complexity research effort at EURISCO. Thank Sébastien. Helen Wilson provided astute advice towards improving the quality of this paper.

References

1. Amalberti, R.: La conduite des systèmes à risques. Le Travail Humain. Presses Universitaires de France, Paris (1996)
2. Aviation Safety Reporting System (ASRS), Background – (24 August 2005) Available from http://asrs.arc.nasa.gov/briefing/br_4.htm
3. Boy, G.A.: Cognitive Function Analysis. Ablex-Greenwood Publishing Group, Westport, CT, USA (1998)
4. Boy, G.A.: Evolution of socio-cognitive models supporting the co-adaptation of people and technology. In: Proceedings of HCI International. New Orleans, USA. (2001)
5. Boy, G.A.: Theories of Human Cognition: To Better Understand the Co-Adaptation of People and Technology, in Knowledge Management, Organizational Intelligence and Learning, and Complexity. In: Kiel, L.D. (ed.) Encyclopedia of Life Support Systems (EOLSS), Developed under the Auspices of the UNESCO, Eolss Publishers, Oxford, UK (2002), http://www.eolss.net
6. Boy, G.A.: Procedural interfaces. In: Proceedings of HIM'02 (the Francophone Conference on Human-Computer Interaction), ACM Press (ACM Digital Library), New York, USA (2002)
7. Boy, G.A.: L'ingénierie cognitive: IHM et Cognition. [Cognitive Engineering: HCI and Cognition] HERMES Sciences, Lavoisier, Paris (2003)
8. Boy, G.A.: Knowledge Management for Product Maturity. In: Proceedings of the International Conference on Knowledge Capture, Banff, Canada, ACM Press, New York (2005)
9. Boy, G.A.: Human-centered design: The AUTOS pyramid. EURISCO International Newsletter #4, Fall (2005)
10. Boy, G.A., De Brito, G.: Toward a Categorization of Factors related to Procedure Following and Situation Awareness. In: Proceedings of the HCI-Aero 2000 Conference. In Cooperation with ACM-SIGCHI, Cepadues, Toulouse, France (2000)

11. Boy, G.A.: Bradshaw, J.R.: Perceived complexity versus internal complexity: Did we take into account expertise, reliability and cognitive stability? In: Proceedings of the Second Symposium on Resilience Engineering. Juan-Les-Pins, France (November 8-10, 2006)
12. Funk, K.: Cockpit Task Management: Preliminary Definitions, Normative Theory, Error Taxonomy, and Design Recommendations. The International Journal of Aviation Psychology 1(4), 271–285 (1991)
13. Funk, K., Chou, C., Madahavan, D.: Preliminary Cockpit Task Management Research at Oregon State University (1999), http://flightdeck.ie.orst.edu/CTM/osuStudies.html
14. Grote, G., Boy, G.A.: Coordination, authority trading, and control in the context of safety-critical systems. In: Proceedings of HCI International 2005, Las Vegas, USA (2005)
15. Hollnagel, E.: The phenotype of erroneous actions: Implications for HCI design. In: Weir, G., Alty, J. (eds.) Human-computer interaction and complex systems, pp. 73–112. Academic Press, London (1991)
16. Miller, G.A.: The magical number seven plus or minus two: some limits on our capacity for processing information. Psychological Review 63 (1956)
17. Nielsen, J.: Usability Engineering. Morgan Kaufmann, San Francisco (1993) ISBN-10: 0125184069
18. Norman, D.A.: Categorization of action slips. Psychological Review 88, 1–15 (1981)
19. Norman, D.A.: Cognitive engineering. In: Norman, D.A., Draper, Steven (eds.) User Centered System Design: New Perspectives on Human-Computer Interaction, Lawrence Erlbaum Associates, Mahwah (1986)
20. Norman, D.A.: The Design of Everyday Things. New York, Doubleday (1988)
21. Norman, D.A.: Complexity versus Difficulty: Where should the Intelligence Be? Plenary Address. In: IUI'02 International Conference on Intelligent User Interfaces Miami, FL, USA (2002)
22. Reason, J.T.: Human Error, Cambridge: Cambridge University Press traduit en français, 1993, Erreur Humaine, Paris, P.U.F (1990)
23. Still, D.L., Temme, L.A.: OZ: A Human-Centered Computing Cockpit Display. Institute of Human and Machine Cognition, University of West Florida, Pensacola, Florida (2003)
24. Thibos, L.N., Bradley, A.: Modeling off-axis vision-II: The effects of spatial filtering and sampling by retinal neurons. In: Peli, E. (ed.) Vision Models for Target Detection and Recognition, pp. 338–379. World Scientific Press, Singapore (1995)
25. Thibos, L.N., Still, D.L., Bradley, A.: Characterization of spatial aliasing and contrast sensitivity in peripheral vision. Vision Research 36, 249–258 (1996)
26. Williams, D.R., Coletta, N.J.: Cone spacing and the visual resolution limit. Journal of the Optical Society of America A4, 1514–1523 (1988)
27. Yates, F.A.: The Art of Memory. In: French translation by Daniel Arasse, 1975, Editions Gallimard, Paris, France (1966)

Computer-Supported Creativity: Evaluation of a Tabletop Mind-Map Application

Stéphanie Buisine[1], Guillaume Besacier[2], Marianne Najm[1], Améziane Aoussat[1], and Frédéric Vernier[2]

[1] ENSAM-LCPI, 151 boulevard de l'Hôpital,
75013 Paris, France
[2] LIMSI-CNRS, BP 133,
91403 Orsay Cedex, France
stephanie.buisine@paris.ensam.fr

Abstract. The aim of this study is to investigate the usability and usefulness of interactive tabletop technologies to support group creativity. We implemented a tabletop interface enabling groups of 4 participants to build mind-maps (a tool for associative thinking). With 24 users in a within-group design, we compared its use to traditional paper-and-pencil mind-mapping sessions. The results showed no difference in idea production, but the tabletop condition significantly improved both subjective and collaborative dimensions, especially by leading to better-balanced contributions from the group members.

Keywords: Creativity, Mind-map, Tabletop device.

1 Creativity in Industrial Applications

Creativity is a high-level cognitive process which has given rise to researches in various fields such as Psychology [4, 16], Engineering [7, 9] or Human-Computer Interaction [3, 6, 14, 15]. Creativity applies to artistic work (e.g. fine arts, literature, architecture, music), educative domain (e.g. early-learning and playing activities), scientific skills (e.g. problem resolution, discoveries, epistemological breakthroughs), and industrial applications (e.g. creation of product functions, stylistic design of artifacts).

In this paper we consider creativity in industrial applications, for example when some people design a product with new innovative functions (e.g. a mobile phone including a positioning system) or search some applications to a new technology (e.g. portable MP3 players). Understanding and supporting this kind of creativity is not only an interesting research challenge: it also bears a strong potential for enhancing industrial innovation and market opportunities.

2 Enhancing Creativity

To improve creativity, a wide-spread practice in companies is the group brainstorming. Although creativity fundamentally remains an individual capacity, it

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 22–31, 2007.

proves to be influenced by the subject's environment: in this respect, collective creativity phenomena are often observed when group emulation improves the expression of one's own creative potential. This is especially true for industrial creativity which can benefit from multiple, or even multidisciplinary viewpoints.

To further improve these collective creativity sessions, methodological toolkits [7, 9] have been formalized to structure the reflection and manage groups' dynamics. Consulting services specialized in creative problem solving also appeared to help companies conduct creativity sessions and apply these methodologies.

Moreover, computer applications have been developed to support industrial creativity[1]. According to Shneiderman [14], the existing software solutions can be categorized into three approaches: inspirational tools (e.g. favoring visualization, free association, or sources of inspiration), structural tools (e.g. databases, simulations, methodical techniques of reasoning), and situational tools (e.g. based on the social context, enabling peer-consultation, or dissemination). Lubart [8] adopted a classification grounded on the role played by the computer in the creative process: systems assisting the user in the management of creative projects (computer as nanny), those supporting communication and collaboration within a team (computer as pen-pal), systems implementing creativity enhancement techniques (computer as coach) and those contributing to the idea production (computer as colleague).

In this context, our goal is to investigate the capacity of a tabletop computer (as a physical device and as a digital interface) to support collaborative creativity related to industrial issues.

3 Tabletop Systems

Tabletop systems (see Fig. 1) are multi-user horizontal interfaces for interactive shared displays. They implement around-the-table interaction metaphors allowing co-located collaboration and face-to-face conversation in a social setting [12, 13]. Tabletop systems are used in various application contexts such as games, photo browsing, map exploration, planning tasks, classification tasks, interactive exhibit medium for museums, drawing, etc. [11]. Such systems being likely to favor collaboration by providing around-the-table visualization facilities, they could be

Fig. 1. Example of a tabletop system using MERL DiamondTouch device [5]

[1] For example Goldfire Innovator (www.invention-machine.com) or ThoughtOffice (www. Idea center.com).

thought of for supporting creativity sessions: in this respect they would fall into both inspirational and situational creativity tools [14] or play pen-pal and coach roles [8] in the creative process. Indeed, some tabletop systems were previously considered for supporting creativity [17, 18] but their actual benefits were not experimentally measured. To assess the usability and usefulness of tabletops in the context of creativity sessions, we believe that it is necessary to compare their use with a control condition relying on traditional paper-and-pencil tools. In the following section, we introduce a creativity application we have developed for tabletop display in order to conduct such an experiment.

4 Our Tabletop Mind-Map Application

4.1 Mind-Maps as a Collective Creativity Tool

In general, creativity methodological framework [7] support a two-step process: first, diverging by producing a vast number of ideas, then converging by selecting a few of them to be further developed. Mind-maps [2] are used in the diverging step. The mind-map principle is based on associative logics and is used for defining the problem to address. The field to explore is written in a central box and the participants express their free associations to this concept. Those ideas are written in new boxes placed as a crown around the central concept (see Fig. 2). A second association level is then built from the primary ideas, etc. Because the second level of association is not directly related to the initial problem, new-original research directions can appear and the realms of possibility grow. Mind-mapping can be performed individually or in a group session. In the latter case, the session has to be managed by an animator whose role is to coordinate speech turns and ensure that the group agrees on every idea. Many software solutions[2] for desktop computers have been developed to support mind-mapping, but none is adapted to tabletop interaction. This is why we implemented our own tool.

Fig. 2. Example mind-map

[2] A complete list of mind-mapping software is available at www.mind-mapping.org, the market leader being MindManager (www.mindjet.com).

4.2 Implementation

Implementation of Tabletop Mind-Maps (TMM, see Fig. 3) was conducted with the DiamondSpin toolkit [13]. TMM was also based on our previous experience with a hierarchy view, namely the Personal Digital Historian (PDH) application [12], which is dedicated to organizing family pictures according to a hierarchy of people and concepts present in the pictures.

A TMM session starts with a root label forced to be in center. The root displays the field to explore, which is important to keep in mind, so we duplicated the label along a symmetry axis to have it more readable from every point of view around the table.

The mind-map is then built top-down when users create new nodes with double-tap-and-drop interaction. This concatenation of double-tap and drag-and-drop appeared to be natural and easy to perform with direct manipulation. The double-tap creates a new node in the sub-hierarchy of the tapped node, while the drag-and-drop specifies the new node position. The background color of the node represents its level in the hierarchy (green for 1st level, blue for 2nd level…).

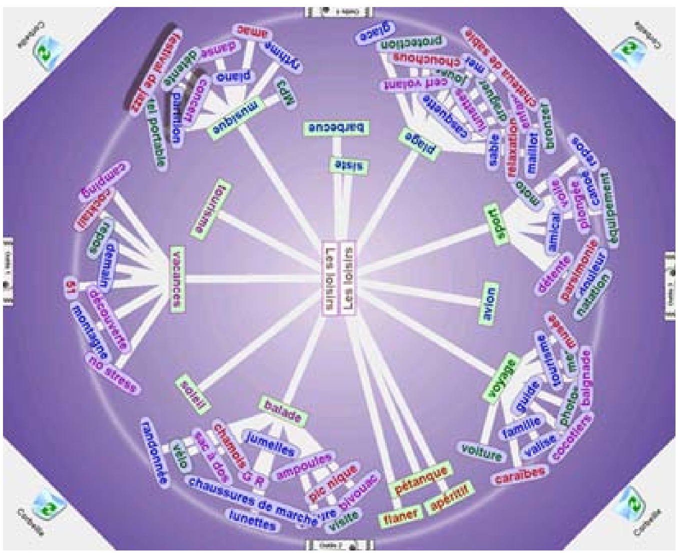


Fig. 3. Example mind-map created with our tabletop application

TMM nodes are editable. The choice of a label being a collaborative activity in mind-mapping, this aspect had to be reproduced in TMM: we chose to allow text input only from a single source, i.e. a physical wireless keyboard with a particular focus management. Indeed, in a tabletop system there can be more than one focused or selected element, as the users interact seamlessly together or in parallel. We made the keyboard focus persistent until the Enter key is pressed. While the text is being keyed, users can create new nodes or select other ones (e.g. to check for possible redundant node's name) without interfering with the edition. The font color of the node represents the user who created it.

Nodes of the hierarchy are freely relocatable on the table. The nodes of a sub-hierarchy will also follow their parent node when the latter is moved on. The orientation of the nodes is adjusted online while they are being moved on so that the text is always oriented outwards to be readable by the nearest user. Moreover, users can rotate the whole display if they want to change the view without changing the arrangement of the hierarchy.

Finally, we introduced a means of creating a temporary view of a sub-hierarchy. A given node becomes the new central root, all the items outside of its sub-hierarchy being temporarily hidden.

5 Experimental Study

This experiment was designed to evaluate the use of a tabletop interactive application for mind-mapping by comparing it with a control paper-and-pencil condition.

5.1 Method

Participants. 6 groups of 4 participants took part in the experiment. Each group included students, professors and/or employees. We excluded groups only composed of students in order to avoid too much familiarity among participants and simulate more realistic conditions of creativity sessions. Overall, users' age ranged from 20 to 52 years old (mean = 28.7; SD = 7.9) and each group was composed of 2 male and 2 female participants.

Materials. For the tabletop condition, we used MERL DiamondTouch [5]: the participants were seated around the table with the experimenter sitting aside on a highchair. The participants interacted on TMM display with finger-input to create, edit or move the mind-map items. The experimenter typed down the content of the nodes using the wireless keyboard.

In the control condition the participants were seated in front of a paperboard with the experimenter standing beside it. The experimenter used a marker pen to build the mind-map and write down its content according to the participants' indications.

Procedure. Each group had to build 2 mind-maps on different topics: 1 in the tabletop condition and 1 in the control condition. The topics were related to the sectors of "Media" and "Leisure": such topics simulate potential reflection for e.g. companies trying to find a way to diversify, searching an application for a new technology or trying to find new markets. These 2 topics were chosen so as to be equivalent in level of abstraction and width of scope. The order of conditions and the assignment of topics were counterbalanced across the whole sample (see Table 1).

Table 1. Counterbalancement of conditions: For each group (A to F), this table defines the order of the 2 conditions (Tabletop and Control) and the topic addressed in each case (Media and Leisure sectors)

Group ID	First mind-map	Second mind-map
A	Tabletop: *Media*	Control: *Leisure*
B	Tabletop: *Leisure*	Control: *Media*
C	Tabletop: *Media*	Control: *Leisure*
D	Control: *Media*	Tabletop: *Leisure*
E	Control: *Leisure*	Tabletop: *Media*
F	Control: *Media*	Tabletop: *Leisure*

To conduct the session, the experimenter first asks the general question "What does *leisure* (resp. *media*) make you think of?" The participants freely suggest some ideas and concepts associated to the target sector, and the experimenter writes down the ideas the group agrees on. Once the first level of the mind-map is completed, the same process is repeated for the second level by focusing on first-level ideas one by one ("What does *xxx* make you think of?"). In this experiment the mind-maps were limited to 2 levels and the time to build them to 10 minutes. The differences between tabletop and control conditions in building the mind-maps are summarized in Table 2.

Table 2. Differences between tabletop and control conditions in the process of mind-mapping

Factor	Description
Spatial position of participants	Around the table vs. in front of the paperboard
Creation of new boxes	By the participants in the tabletop condition vs. by the experimenter in the control condition
Modification / suppression of a box	Allowed in tabletop but not in control condition
Spatial arrangement of items	Online modifications allowed in tabletop but not in control condition
Rotation of the mind-map	Allowed in tabletop but not in control condition
Focus on a first-level idea	Explicit in tabletop (making the rest of the mind-map disappear) vs. implicit in control condition (whole map always displayed)

The tabletop condition was preceded by a familiarization phase for demonstrating the table's functionalities to the participants. Both tabletop and control conditions were then video-recorded. At the end of the experiment, users had to fill in a questionnaire to assess the following dimensions: efficiency, usability, usefulness of the tabletop system, satisfaction, and comparison with the control condition. Users had to quantify their impressions on 7-point scales and were particularly prompted to complete with free qualitative comments. The whole experiment lasted about 1 hour for each group.

5.2 Data Analysis

Inferential analyses were performed by means of ANOVAs using SPSS software. Three dimensions were investigated: the performance in mind-mapping, the subjective experience of users, and the collaborative behaviors.

Performance. We chose to assess the performance dimension from the exhaustiveness of the outcome. As we lack absolute standards to evaluate a mind-map in itself, we decided to aggregate the mind-maps of the 6 groups for the same topic and take this as a reference to be compared to each mind-map. We rated the exhaustiveness of the mind-maps by considering both the total number of ideas and the number of categories of ideas in comparison to the reference.

Subjective experience. This dimension was computed from the questionnaire ratings. The analysis processed on these data also accounted for users' gender and category (student, professor or employee).

Collaboration. The participants' collaborative behaviors were annotated from the video-recordings of the sessions. We collected the following behaviors: assertions (e.g. giving an idea), information requests, action requests, answers to questions, expression of opinions, communicative gestures related to the task, and off-task talks. The "communicative gestures" variable includes e.g. pointing to the map, interrupting s.o. or requesting the speech turn by a gesture, which can be observed in both conditions. In the tabletop condition, it also includes gesture-inputs on the table, with the exclusion of creation / edition / suppression actions which we did not consider as communicative gestures.

We first analyzed the raw behavioral data for each participant, and then we converted them into percentages in order to assess the respective contribution of each participant in the group. Such an index finally enabled us to compute the difference between the actual collaboration pattern of each group and a theoretical perfectly-balanced pattern (each one of the 4 participants would contribute 25%).

5.3 Results

Performance. No significant difference appeared between tabletop and control conditions on our index of exhaustiveness of mind-maps ($F(1/5) = 0.92$, NS).

Subjective experience. There was no significant effect of the condition (tabletop vs. control) on easiness ($F(1/20) < 0.1$, NS) and efficiency ($F(1/20) = 1.02$, NS) of mind-map building. However, the tabletop was rated as significantly more pleasant to use ($F(1/20) = 10.43$, $p = 0.004$), enabling a more pleasant communication between participants ($F(1/20) = 5.01$, $p = 0.037$), more efficient group work ($F(1/20) = 3.56$, $p = 0.074$) and more pleasant group work ($F(1/20) = 4.23$, $p = 0.053$) – see Fig. 4. Users' gender and category had no influence on any of the previous results.

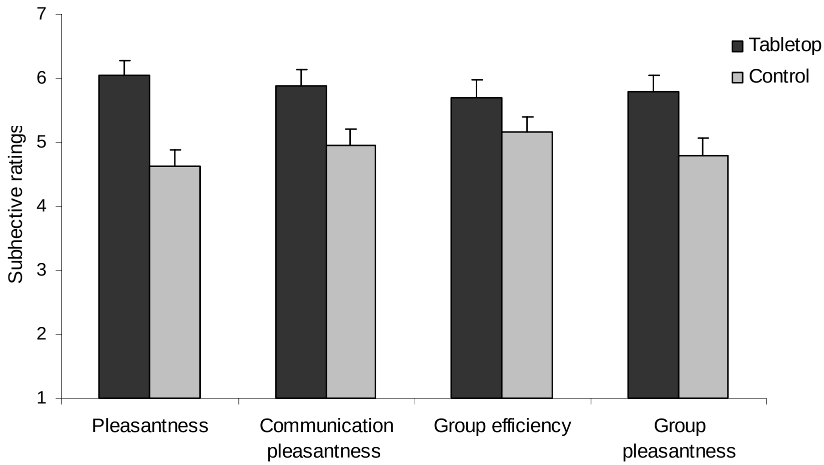


Fig. 4. Subjective ratings of participants for the tabletop and control conditions

Collaboration. The variables "expression of opinion" and "off-task talk" comprised too many missing values to be analyzed. The other raw behavioral data showed no significant difference in the absolute number of any of the variables, except for the communicative gestures category: tabletop led to more communicative gestures than control condition (F(1/22) = 3.59, p = 0.071).

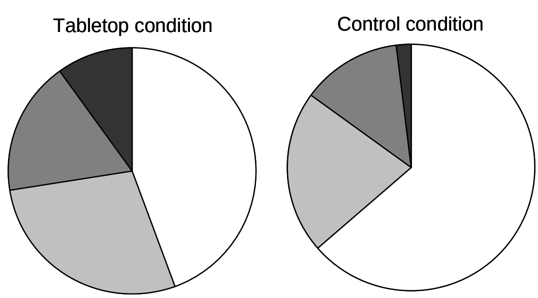


Fig. 5. Collaboration patterns in tabletop (left) and control (right) conditions: this graph represents the average contribution of the 4 participants ranked on a leader / follower scale. This figure illustrates that the contributions of the participants were significantly better balanced in the tabletop condition (p = 0.013).

The analysis of collaboration patterns showed that participants' verbal contributions (sum of all behaviors without communicative gestures) were significantly better balanced in tabletop than in control condition (F(1/22) = 7.35, p = 0.013) – i.e. they were significantly closer to the theoretical perfectly-balanced pattern. Fig. 5 presents the average collaboration patterns in both conditions: to obtain this figure, we ranked the participants of each group from the one who contributed the

most (the leader) to the one who contributed the least (the follower) and averaged the data for the 6 groups. The same result applies for communicative gestures: the gestural contributions were significantly better balanced in tabletop than in control condition ($F(1/22) = 8.94$, $p = 0.007$).

6 Conclusion and Future Work

The tabletop condition significantly improved both subjective and collaborative dimensions of mind-mapping. First of all, the participants found that the tabletop system was more pleasant to use, improved group communication and collaboration efficiency. These effects on users' impressions could be explained e.g. by the spatial position of participants favoring social interaction, the attraction of a new technology, and/or the more active involvement of participants in this condition.

Moreover, the behavioral analysis showed that the tabletop system enabled a better collaboration: while the control condition showed strong leaders and followers, in the tabletop condition the participants collaborated in a better-balanced way. Some benefits of a tabletop system compared to a wall display or a desktop computer were previously observed by Rogers and Lindley [10] but their setting was noticeably different from ours: their tabletop device supported only single-touch interaction (with a pen) and a single viewpoint (so that the participants had to sit side by side and not around the table). They observed more interaction and role changing (swapping the possession of the input device) in the tabletop condition: it proved easier and more natural to change roles because of the use of a direct input device (a pen has to be placed directly on the display whereas a mouse controls the pointer from a distance) and because of the physical proximity of the participants to this input device (higher in the tabletop than in the wall display condition). In our experiment the collaborative benefits cannot be explained by any of these reasons because all 4 participants had the same role and interaction capacity. We could tentatively explain our results by the spatial position of people around the table, which can facilitate idea exchange, or by the attraction of a new technology, which could prompt the participants to interact with the tabletop interface and thus to give new ideas. The second hypothesis is less likely because it may have resulted in higher performance in idea generation. Therefore we hypothesize that the collaborative benefits we observed come from the around-the-table placement of people. This assumption will be tested with a new control condition where the participants will have to build a paper-and-pencil mind-map around a table. This new experiment will also complete the data about the subjective preferences expressed in the present study.

Finally, despite all the advantages of our tabletop application (subjective engagement, better collaboration, active involvement of users, focus on first-level ideas, flexibility of the mind-map display…), the experimental results showed no significant difference in the quality of outcomes between tabletop and control conditions. In the next steps of the project, we intend to focus more deeply on the performance dimension and search a way to improve it. We should develop a more accurate analysis of mind-map process and outcome to better understand the idea production mechanism. We also plan to test the influence of innovative interaction styles (see e.g. the paper metaphor [1]) on idea production and organization.

The global experimental process followed in this study (comparison of tabletop and traditional paper-and-pencil condition, variables collected…) is currently being applied to other creativity tools such as brainstorming on sticky notes in order to investigate whether the present results apply to other situations of group creativity.

Acknowledgements. This study was supported by the ANR-RNTL DigiTable project (www.digitable.fr).

References

1. Besacier, G., Rey, G., Najm, M., Buisine, S., Vernier, F.: Paper metaphor for tabletop interaction design. In: HCII'07 Human Computer Interaction International. LNCS, Springer, Heidelberg (2007)
2. Buzan, T.: The Mind Map Book. Penguin Books (1991)
3. Candy, L., Hori, K.: The digital muse: HCI in support of creativity. Interactions 10, 44–54 (2003)
4. Csikszentmihalyi, M.: Creativity: Flow and the psychology of discovery and invention. Harper Perennial, New York (1996)
5. Dietz, P.H., Leigh, D.: DiamondTouch: A multi-user touch technology. In: UIST'01 International Conference on User Interface Software and Technology, pp. 219–226. ACM Press, New York (2001)
6. Farooq, U.: Eureka! Past, present, and future of creativity research in HCI. ACM Crossroads 12, 6–11 (2005)
7. Isaksen, S.G., Dorval, K.B., Treffinger, D.J.: Creative approaches to problem solving: A framework for change. Kendall Hunt (2000)
8. Lubart, T.: How can computers be partners in the creative process. International Journal of Human-Computer Studies 63, 365–369 (2005)
9. Osborn, A.F.: Applied Imagination. Principles and procedures of creative problem-solving. Charles Scribner's Sons (1953)
10. Rogers, Y., Lindley, S.: Collaborating around vertical and horizontal large interactive displays: Which way is best? Interacting with Computers 16, 1133–1152 (2004)
11. Scott, S.D., Carpendale, S. (eds.): Interacting with digital tabletops. Special issue of IEEE Computer Graphics and Applications, 26 (2006)
12. Shen, C., Lesh, N., Vernier, F.: Personal Digital Historian: Story sharing around the table. Interactions 10, 15–22 (2003)
13. Shen, C., Vernier, F., Forlines, C., Ringel, M.: DiamondSpin: An extensible toolkit for around-the-table interaction. In: CHI'04 International Conference on Human Factors in Computing Systems, pp. 167–174. ACM Press, New York (2004)
14. Shneiderman, B.: Creating creativity: User interfaces for supporting innovation. ACM Transactions On Computer-Human Interaction (TOCHI) 7, 114–138 (2000)
15. Shneiderman, B., Fischer, G., Czerwinski, M., Resnick, M., Myers, B.: Creativity support tools: Report from a U.S. National Science Foundation sponsored workshop. International Journal of Human-Computer Interaction 20, 61–77 (2006)
16. Sternberg, R.J.: Handbook of Creativity. Cambridge University Press, Cambridge (1998)
17. Streitz, N.A., Geißler, J., Holmer, T., Konomi, S.: I-LAND, an interactive landscape for creativity and innovation. In: CHI'99 International Conference on Human Factors in Computing Systems, pp. 120–127. ACM Press, New York (1999)
18. Warr, A., O'Neill, E.: Public Social Private Design (PSPD). In: CHI'06 International Conference on Human Factors in Computing Systems, pp. 1499–1504. ACM Press, New York (2006)

Symbiosis: Creativity with Affective Response

Ming-Luen Chang and Ji-Hyun Lee

Graduate School of Computational Design,
National Yunlin University of Science & Technology,
123 Section 3,
University Road, Douliou,
Yunlin 64002, Taiwan, R.O.C
{g9434703, jihyun}@yuntech.edu.tw

Abstract. The objective of this research is to present the symbiosis concept that integrates creativity and the recent research issues in affective response to products shapes. The major idea behind this study is systematically using affective response and design axiomatic in rational way through creativity approach that support on creativity stimulation for current highly competitive market. The practicality of the proposed methodology involved affective response measurable system that based on Semantic Differential (SD) method and interrelated computational regulation, creativity approach that based on Sensuous Association Method (SAM) and Creativity-Based Design Process (CBDP), and integrated mechanism using Axiomatic Design (AD) method.

Keywords: Affective Response, Creativity Approach, Axiomatic Design Method.

1 Introduction

In the highly competitive market, the customer-oriented creativity has become a great concern of most companies [3, 6, 13, 19] as shown in Fig. 1. How to conduct customers' affection into the process of product shape manipulation is a new trend and strategy, which called "Form follows Affection".

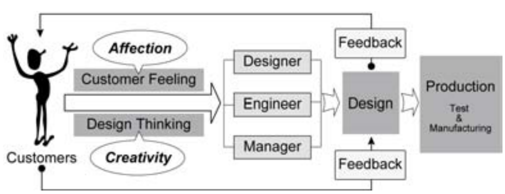


Fig. 1. Diagram of the customer-oriented New Product Design (NPD) process concept

Affective response is said to be a common customers' psychological response to the perceptual design details of the product [7, 15]. When customers contact with a specific product, the shape can evoke specific affection. A growing number of

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 32–41, 2007.

research studies are now available to shed some light on relationship perception among affective response and product shape [1, 2, 4, 9, 10, 11, 13, 19]. However, few studies have been down on the effect of how make the creativity product which involve affective response. According to above researches, how creativity stimulating to designers that based on the affective response considered into design thinking field become the new design issue. The purpose of this paper, therefore, is to present the symbiosis concept that integrates creativity and the recent research issues in affective response to products shapes.

In the view of the above research purpose, several major sets of research points to be addressed in this study are as follows: (1) how affective response can be acquired and measuring, (2) how affective response can translating into creativity thinking, (3) what approach can encourage creativity, (4) how achieved symbiosis concept that mechanisms can support each other. To achieve those research points, several tasks are structured as follows. The second section deals with the theoretical foundations on affective response and creativity for the development of the research. After that, research methodology is presented, with affective response measurable system, creativity approach and integrated mechanism. System architecture and a set of operating interfaces are described for the research implementation. The mobile phone as the case study used in the implementation.

2 Background Review

2.1 Affective Response on Product Features

2.1.1 Customers' Affective on Product Shapes

Crozier (1994) indicated that the psychological responses to products are influenced by the product's appearance and that's why product appearance plays a significant role that could convey a designer's ideas and provide consumer visual references in affective response. Customers who are inexperienced with a product may focus primarily on the first impression and the styling of the product; they expect a product to be a living object that expresses an emotional image via its shape [10, 15]. Therefore, customers' affective of product shape become an important issue for designers and highly competitive market strategy.

2.1.2 The Measurement of Customer's Affective Response

Subjective assessments are commonly used to evaluate affective response; ask persons and they will answer how they feel and what they like. It is, however, important to conduct such assessments in a structured manner so that the results are reliable and valid and can be compared across different products and different cultures. In order to investigate the customer's perception, feeling and emotion, Osgood et al. (1957) propose Semantic Differential (SD) method, which is one of the most frequently used procedures for getting meaning space from well-prepared samples by investigation of the qualitative scale using numerical mapping relationship between the samples and the related words and convert into proper numerical data [13, 15, 21]. In this method, the subject's perception of product forms is quantified on a numerical scale. Many researchers have used this method to study

specific aspects of product form, including styles, colors, and other attributes in product design [11].

2.2 Creativity Approach on Design Process

2.2.1 Sensuous Association Method (SAM)

The sensuous association method (SAM) is developed by Chou et al. (2004). This creativity method is based on the naturally sensuous ability of human being that used for refreshment of sensuousness and association of inspiration, and it can be regarded as a creativity tool for encouraging designers' potential to produce innovative ideas quickly. The method contains four personal behaviors of human sensuousness and one extrinsic influence of the environment. They are expressed as follows (Fig. 2):

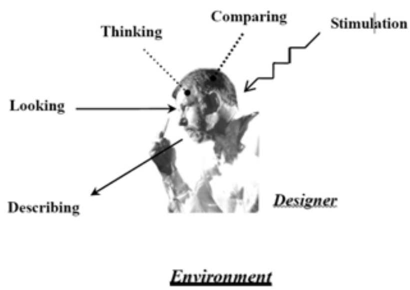


Fig. 2. Diagram of the Sensuous Association Method (SAM) (from [3])

(1) *Looking*: information input course, (2) *Thinking*: inference and re-association course. (3) *Comparing*: extraction and restructuring course. (4) *Describing*: creativity output course. (5) *Stimulation*: catalysis and outburst course.

2.2.2 Creativity-Based Design Process (CBDP)

Product development is often described as an iterative process to find solutions that fulfill a given requirement specification [17]. Jones (1992) proposes design process includes three essential stages: (1) Divergence, (2) Transformation, and (3) Convergence. As shown in Fig. 3., the *divergent* stage is an analytic process for searching the problem space, which can be described as "breaking the design problem into pieces". The *transformation* stage is a synthetic process for generating the solution space, characterized as "putting the pieces together in new ways". The *convergent* stage is an integration and evaluation process for finding applicable sub-solutions and optimal design solutions, described as "testing to discover the results of putting the new arrangement into practice".

2.3 Axiomatic Design (AD) Method

To design, we have to go from "what" in the functional domain to "how" in the physical domain, which requires mapping. Axiomatic design is a principle-based

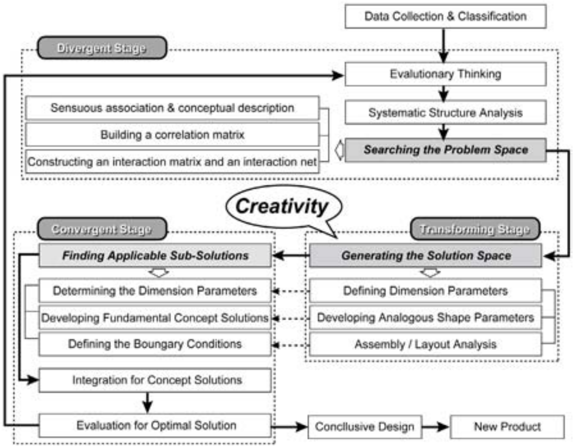


Fig. 3. Framework of the Creativity-Based Design Process (Re-drawn from [3])

design method focused on the concept of domains guides us to mapping among design requirement, design solution and decomposition developed by Dr. N.P. Suh at MIT [16]. The primary goal of axiomatic design is to establish a systematic foundation for design activity by two fundamental axioms. The basic postulate of the axiomatic approach to design is that there are fundamental axioms that govern the design process. The axioms are formally stated as:

Axiom 1: The Independence Axiom - Maintain the independence of Functional Requirements (FRs).
Axiom 2: The Information Axiom - Minimize the information content in design.

The first axiom is called the independence axiom focuses on the nature of the mapping between "what is required" (FRs) and "how to achieve it" (DPs). It states that a good design maintains the independence of the functional requirements, where FRs are defined as the minimum set of independent requirements that characterize the design goals. The second axiom is called the information axiom establishes information content as a relative measure for evaluating and comparing alternative solutions that satisfy the independence axiom where the design that has the smallest information content is the best design [16, 20].

3 Methodology

3.1 Affective Response Measurement

3.1.1 Knowledge Acquisition Using SD Method

SD method is frequently used to acquire and evaluate the customers' affective response. The historical data about the customers' affective needs of mobile phones assorted according to well-known affective words related to mobile phones includes: Portable, Sturdy, Enjoyable, Dignified, Cheerful, Natural, Delightful, Stimulating, Comfortable, Dazzling, Mature, Fashionable, Friendly, Cute and Futuristic [13, 15]. The affective words will be store into affective database.

3.1.2 The Measurement Mechanism

Based on SD method, Yang et al. (1999) propose the mathematics equations for SD method measurement. The sequential scaled numbers provide quantitative information for each affective word, it define S_{jk}^{λ} as SD method scores for subjects, where λ is design element number, i is affective word number, and j is item number. Also can define SM_i^{λ} as average value for each word on each element, shown in equation (1), where n is the total number of subjects, which means a representative value of the target customer group for each affective word on each design element. These values for each sample are used as criterion variables in estimating the relationship between affective words and design elements in the regression model. Equation (2) is to find coefficients a_{jk} in order to minimize the deviation between estimated values and real values.

$$SM_i^{\lambda} = \sum_{j=1}^{n} S_{ij}^{\lambda} \tag{1}$$

$$y^{\lambda} = \sum_{j=1}^{m}\sum_{k=1}^{c_j} a_{jk} x_{jk}^{\lambda} + e^{\lambda} \tag{2}$$

$$x_{jk}^{\lambda} = \begin{cases} 1,\text{where a sample } \lambda \text{ corresponds to item j and category k,} \\ 0, \text{ otherwise} \end{cases} \qquad \sum_{k=1}^{Cj} x_{jk}^{\lambda} = 1. \tag{3}$$

a_{jk} is called partial regression coefficients or category score (weight). y^2 and x_{jk}^{λ} is the criterion variables and explanatory variables, respectively. The estimated values of a_{jk} can be derived by solving a simultaneous equation composed of equation (2). For example, if there are fifty samples, fifty simultaneous equations can be composed. Practically, criterion variables y^{λ} correspond to SM_i^{λ} gained from SD evaluation and explanatory variables x_{jk}^{λ} have 0 or 1 according to the composition of the design elements on each sample [21].

3.2 Creativity Approach That Integrates SAM and CBDP

According to the foundation theories of creativity on behaviors of SAM and process of CBDP, the synthetic model for describe creativity approach shown in Fig. 4.

The Divergent can occur among Thinking stage to Describing stage; after describing, through Comparing and re-Thinking, than re-Describing can be Transforming stage and stimulation; the recursion from describing stage to comparing stage can be Divergent.

3.3 Integrating Axiomatic Design (AD) Method to the Creativity Approach

In order to achieve the integration mechanism, the AD method is the way it supports on combination work. The integration diagram is shown in Fig. 5. Based on AD method theory, the mapping process between the domains can be expressed mathematically in terms of the characteristic vectors that define the design goals and design solutions (Fig. 5).

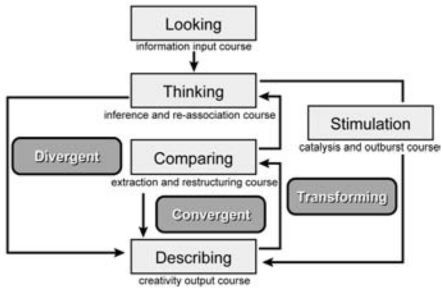


Fig. 4. Diagram of Creativity Approach

To design goals, dependent on Independence Axiom of axiomatic design theory, which classifies the design into three A matrix: Uncouple design, Decouple design and Couple design as shown in equation (4) [12]. Supporting that A matrix into process of the creativity mechanism. Decouple matrix support on process of the Divergent to Transforming; Couple matrix support on process of the Transforming to Convergent; and Uncouple matrix support on process of the Convergent to Divergent.

$$\begin{pmatrix} x & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & x \end{pmatrix} \quad \begin{pmatrix} x & 0 & 0 \\ x & x & 0 \\ x & x & x \end{pmatrix} \quad \begin{pmatrix} x & x & x \\ x & x & x \\ x & x & x \end{pmatrix} \tag{4}$$

Uncouple Decouple Couple

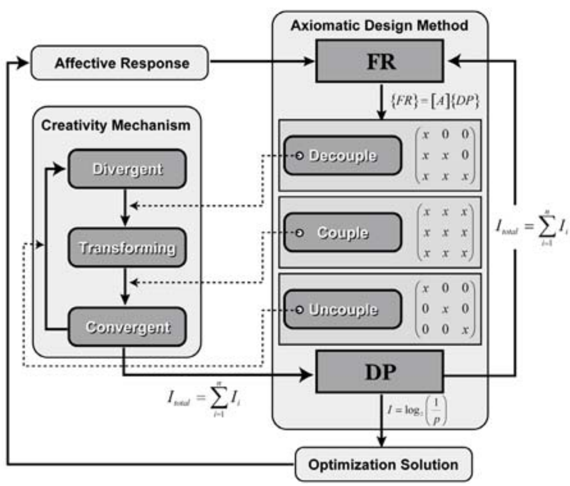


Fig. 5. Diagram of Symbiosis System

Equation (5) is a design equation for the design of a product. The set of functional requirements that defines the specific design goals constitutes the $\{FR\}$ vector in the functional domain. The set of parameters in the physical domain that has been chosen to satisfy the FRs constitutes the $\{DP\}$ vector. $[A]$ is the design matrix that relates FRs to DPs and characterizes the product design.

$$\{FR\} = [A]\{DP\} \tag{5}$$

According to the Information Content Axiom of axiomatic design theory, I_i for a given FR*i* is defined in terms of the probability P_i of satisfying Fr*i* (6). A design's information content is calculated according to the logarithmic expression equation (7). The minimum Information Content (I) is optimization solution. When there are n functional requirements, the total information "content (I_{total})" is given by equation (8).

$$I_i = \log_2 \frac{1}{P_i} = -\log_2 P_i \tag{6}$$

$$I = \log_2 \left(\frac{1}{p} \right) \tag{7}$$

$$I_{total} = \sum_{i=1}^{n} I_i \tag{8}$$

4 Prototype System Implementation

Our system is divided into five parts, which include client graphical user interface (GUI), intra-system GUI for design work stakeholders, affective response measure

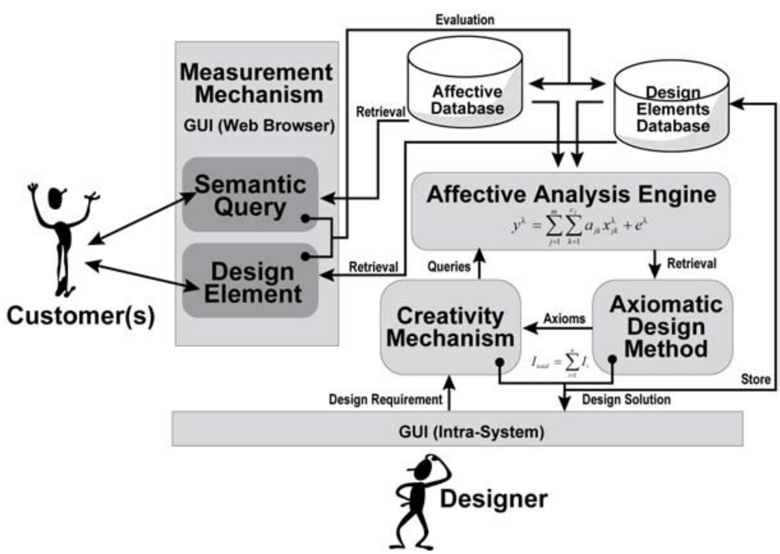


Fig. 6. System Architecture

mechanism, creativity mechanism and AD method axiomatic based. Fig. 6. shows the architecture of system of this paper.

When designers want to create a product, they should know the customers' feeling. The affective response measure mechanism provides to evaluate affective response through semantic query. Customer can see the design elements directly and answer their response via web browser and evaluative data will be stored into affective database automatically. Then, when designers need to call the affective response data, they can directly use intra-system to acquire the information. When designers operate the intra-system, they can input the design requirement that is related with affective response and the system will call data from affective database through affective analysis system and provide the data and statistics into creativity mechanism. Based on the data, designers not only can know the design information depending on the customers' affective response, but also can utilize the data to re-associate, restructure, break or compose, which are supported from axiomatic design (AD) method. During this process, designer can saw the design elements via requirement set, looking in such elements, think about that and they will describe what they think compare with the requirement set. They also can see the description of each result on message area. Of course they can add some comments inside. According their thinking and looking, after comparing among each design elements, they can choose the design element into re-associate area, and then construct the solution, re-analyze with new requirement or break again they can decide. Such process can stimulate the creativity thinking via continuously looking, thinking, describing and comparing. After that, they can using design element to re-associate, re-construct to encourage design think. Finally, through the AD method can support designer to find the optimization solution that conform to requirement. The interface of operating system is shown in Fig. 7.

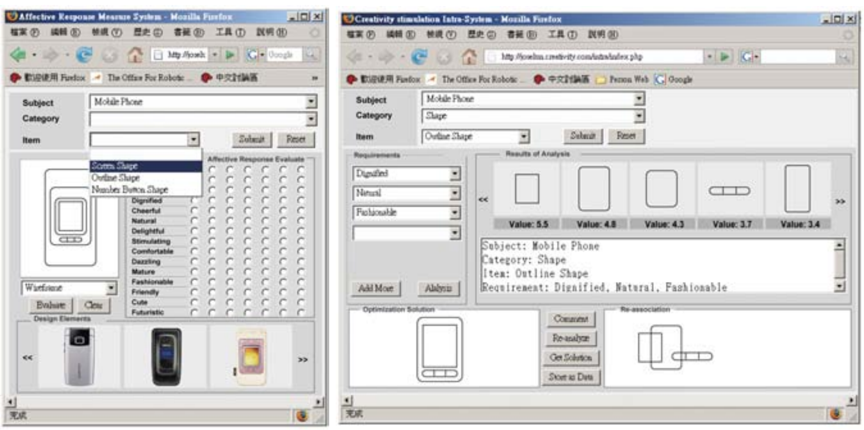


Fig. 7. A snapshot prototype interface of the affective response measure system (client) for customers and creativity stimulation system (intra-system) for design work stakeholders

5 Conclusion

In this paper, we present symbiosis concept that integrates creativity and affective response to product shape as the mentioned at the beginning. Based on several foundational theoretical backgrounds, this paper is systematically using affective response and design axiomatic in a rational way through creativity approach. The prototypical computational tool can encourage creativity thinking to design-related stakeholders through continuous manipulation of convergence-divergence for design elements based on previous case that was evaluated form customers and AD method. In the future, we anticipate that the research is needed on the results of effectiveness of using AD method in creativity process. We are hopeful that future research will provide more detailed experiments for possibility to practical issues.

References

1. Barnes, C., Southee, C., Henson, B.: The Impact of Affective Design of Product Packaging Upon Consumer Purchase Decisions. In: Proceedings of the International Conference on Designing Pleasurable Products and Interfaces, pp. 134-135 (2003)
2. Chen, L.L., Wang, G.F., Hsiao, K.A., Liang, J.: Affective Product Shapes through Image Morphing, In: Proceedings of the International Conference on Designing Pleasurable Products and Interfaces, pp. 11–16 (2003)
3. Chou, J.R.: Applying a Creativity-based Design Process to New Product Development. Ph.D. Dissertation. In: Institute of Industrial Design, National Cheng Kung University, Tainan, Taiwan, R.O.C (2004)
4. Chuang, M.C., Chang, C.C., Hsu, S.H.: Perceptual Factors Underlying User Preferences Toward Product Form of Mobile Phones. Int. J. Industrial Ergonomics 27, 247–258 (2001)
5. Crozier, R.: Manufactured Pleasures-Psychological Responses to Design. Manchester University Press, Manchester (1994)
6. Cross, N.: Engineering design methods: Strategies for product design, 3rd edn. Wiley, Chichester, UK (2000)
7. Demirbilek, O., Sener, B.: Product Design, Semantics and Emotional Response. Ergonomics 46, 1346–1360 (2003)
8. Heap, J.: The Management of Innovation and Design. Cassell, London (1989)
9. Hsiao, K.A.: Affective Responses of Product Shapes. In: Ph.D. Dissertation, Institute of Design, National Taiwan University of Science and Technology, Taipei, Taiwan, R.O.C (2006)
10. Hsiao, K.A., Chen, L.L.: Fundamental Dimensions of Affective Responses to Product Shapes. Int. J. Industrial Ergonomics 36, 553–564 (2006)
11. Hsu, S.H., Chuang, M.C., Chang, C.C.: A semantic Differential Study of Designers' and Users' Product Form Perception. International Journal of Industrial Ergonomics 25, 375–391 (2000)
12. Hu, M., Yang, K., Taguchi, S.: Enhancing Robust Design with the Aid of TRIZ and Axiomatic Design (Part I). TRIZ Journal (2000)
13. Jiao, J., Zhang, Y., Helander, M.: A Kansei mining system for affective design. Expert Systems with Applications 30, 658–673 (2006)
14. John, J.C.: Design Method. Van Nostrand Reinhold, New York (1992)

15. Khalid, H.M., Helander, M.G.: A Framework for Affective Customer Needs in Product Design. Theoretical Issues in Ergonomics Science 5, 27–42 (2004)
16. Lee, D.G., Suh, N.P.: Axiomatic Design and Fabrication of Composite Structures. Oxford University Press, New York (2006)
17. Lee, Y.C., Deng, Y.S.: A Design System Integrating TRIZ Method and Case-Based Reasoning Approach. In: Proceedings of the 8th International DDSS Conference, pp. 387–402 (2006)
18. Osgood, C.E., Suci, G.J., Tannenbaum, P.H.: The Measurement of Meaning. Illinois Press, US (1957)
19. Petiot, J.-F., Yannou, B.: Measuring Consumer Perceptions for a Better Comprehension, Specification and Assessment of Product Semantics. International Journal of Industrial Ergonomics 33, 507–525 (2004)
20. Yang, K., Zhang, H.: A Comparison of TRIZ and Axiomatic Design. In: Proceeding of the 1st ICAD, Cambridge, MA, pp. 235–242 (2000)
21. Yang, S., Nagamachi, M., Lee, S.: Rule-Based Inference Model for the Kansei Engineering System. International Journal of Industrial Ergonomics 24, 459–471 (1999)

Embodied Virtual Agents: An Affective and Attitudinal Approach of the Effects on Man-Machine Stickiness in a Product/Service Discovery

Pablo Lambert de Diesbach[1] and David F. Midgley[2]

[1] ESSEC Asian Center, N.Library13-02, 100 Victoria Str., 188064, Singapore -Sup de Co Group, 17000 La Rochelle, France
diesbach@esc-larochelle.fr
[2] INSEAD, Bd de Constance, 77305 Fontainebleau, France
diesbach@essec.fr, david.midgley@insead.edu

Abstract. Of the objective of this paper is to develop and test a model of the effects of an embodied virtual agent (EVA) on the user of an online interface. The tested interface is a brand website—a possible channel of purchase, but also a media of information about products or services. The process of relationship building between website and user is the focus of interest here, a perspective that is richer than what is often called “acceptability” in the literature. Instead the paper proposes a construct of “stickiness;” i.e. the capacity of the interface to retain the user and to create positive behavioral intentions towards it. An integrative model is proposed. The effects of the presence of an EVA and of its congruency with the website are measured, and two possible routes of influence to stickiness investigated. Simple effects (with no route of influence) are observed on behavioral stickiness, whereas other effects via attitudinal and via affective routes, are observed, on intentional stickiness.

Keywords: embodied virtual agent, attitude, affect, stickiness, relationship, congruency.

1 Introduction

Every year, embodied virtual agents are more widely used on electronic interfaces, including online platforms. Research shows that virtual agents impact human behavior, but most research does not relate this impact to a real, substantive theoretical framework. Yet there are many relevant psychological constructs such as attitudes or affective reactions that may help understand the observed effects. This paper aim to model the influence of an embodied agent and to measure the possible effects of one important construct: namely the agent’s perceived congruency with the site and the brand.

2 Theoretical Framework

2.1 From Selling to Creating a Relationship Online

The internet has become a major vehicle for trade and commerce. Indeed, a number of marketing studies justify studying the internet because of the volume of sales realized online. However, such a view has three limitations.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 42–51, 2007.

First the concept of electronic commerce and internet-based commerce is rarely defined and they do differ. The example of the French internet market and of the stock exchange markets, are particularly relevant. Second, the concept of “online purchase” is very reductive: a number of customers do not exactly use internet to “buy”, rather they use the internet in crucial steps of their information search sequence before buying offline [27, 37, 49, 56]. The data collected and available on the FEVAD (French federation of distant sales) also confirms this. Third, it appears that a crucial step in preparing some possible purchase lies in the capacity of the brand or outlet to create a positive attitude, and hence, predisposition, in the potential consumer: a website may therefore be used not only as an informative tool [32, 33, 34, 35], but also as a relationship builder, a service encounter [24]and an experiential tool [23, 56].

Thus a brand web site may have a socializing capacity, particularly if it displays an embodied agent [47, 48]. In this view, interface humanization may play a crucial role, due to the positive or negative emotions it may generate, and to the possible subsequent effects in term of attitude, intentions and behaviors [41, 51, 52, 53].

2.2 Embodied Virtual Agents

Definition of a virtual agent

A virtual agent is a piece of software that can do some task on behalf of an interface user, or help such user doing this task more efficiently and/or rapidly. It is called “agent” because it acts, i.e. it performs some tasks, and it is said “virtual” because it gives the impression of being “alive”, i.e. it shows some characteristics of a living objects, albeit in a virtual world. Previous research often does not define such virtual world, which is a representation of the real world, mediated through an electronic interface. For example, we do not speak of a virtual world for an interaction on the phone or through the TV or a cinema screen. Why? It actually seems that we need the world to be represented on a close-to-us interface, and one that presents various sensorial modalities. That is, it uses animated images, sound, if possible 3D representations and so forth. It may even use in some cases odour and tactile senses. In common situations, animated images and sound are enough; but still, TV or movies are not considered as “virtual worlds”. It seems that the crucial difference is that a virtual world must offer some degree of interaction and some physical proximity to the user. That is what distinguishes between a simple electronic display and a virtual world. In a virtual world the user feels close-to and immersed in the interface.

Intelligence of an agent

One key characteristics of an agent is its “intelligence”, or capacity to be autonomous and to react in a number of situations, which may have not be totally imagined and integrated into a script by its conceivers. Its intelligence makes the agent more likely to be perceived as credible or natural, precisely because it reacts in a way which makes the user feel the agent could be alive. Yet, credibility or naturalness are important drivers of what is called the acceptability of an agent – even if such construct of “acceptability” [25, 26] may have not been conceptualized with a strict enough meaning.

Embodiment of an agent, definition of an EVA
An agent may be embodied, or not: as a piece of software per se it may not have a visible interface: such interface makes it possible for the user to actually “see” the agent performing a task or interacting, with other agents or with the user herself. At this stage we now can adopt the present definition of an embodied, virtual agent: one proposed by [Burgoon et al, p.554). “Computer interfaces that come in a variety of guises and that present and process information according to a set of predefined algorithms. Agents may be designed to appear more anthropomorphic by fitting them with distinctly human-like (virtual) features such as voice recognition, synthesized voices, and computer animation that simulates human facial expression and gestures.”

2.3 The Concept of Stickiness

A managerial literature exists on the concept of online stickiness, which generally expresses the capacity of a website to retain a user [36, 49]. It was somewhat unquestioned and accepted as a concept of great interest by practitioners, particularly as the driver of a successful branding or commercial website. Unquestioned that is until the internet bubble burst and a necessary paradigm shift became necessary (16, 56]. Mere behavioral retention may be the center of early thinking about internet strategy for a website [13] but it is likely too narrow a view. [23] posits that a more useful construct of stickiness should be composed of behavioral and attitudinal or intentional measurement. This would capture a real desire to maintain a relationship with the interface, and one that could be distinguished from retention driven by mistakes, by poor site ergonomics or by tricky website retention strategies. A more useful construct is therefore defined as a “power of retention”, i.e. of creating a durable relationship with a user [23]. It is operationalised in a single, bi-dimensional construct, composed of a behavioral dimension measured by navigation duration and the number of pages visited, and an intentional dimension measured by both intention to recommend to others and intention to revisit. We posit that the interface stickiness proposed hereafter is also actually a possible – if not exhaustive – measure of the agent acceptability in the following way: agent acceptability would be positively related to the interface (here, the website) acceptability, and interface stickiness is an operationalisation of this interface acceptability.

2.4 Trying to Explain the Effects Via Attitudes and Affect

The construct of attitude
Attitude may be defined as a durable disposition into answering in a constant way to some situation, stimuli, aspects, characteristics, of an object. The referred “object” may be a person, an environment, and also a brand, a product or service, an outlet [3, 4, 6, 18, 19]. Attitudes would explain intentions and behaviors. The attitudinal approach has been criticized and enriched in a number or works [5] as being too rationalistic and not encompassing the affective dimension adequately. Second, the concepts of attitude and the concept of hierarchy of effects of attitudes have been extended to the internet context [12, 14, 55]. It could explain the effects of an agent on variables such as behaviors and intentions.

The relevance and limitations, of the Environmental psychological approach
A stream of research has integrated the environmental psychology framework [43, 44] into modeling the effects of ambience and design factors on reactions, intentions, behaviors, in a behaviorist and then a modified, enriched approach encompassing the chain stimulus-emotional responses (called “organism”, hence the name SOR)-behavioral responses. Such approach is adopted in marketing and it partly explains behaviors when in an outlet [28, 29, 30, 31, 54], in an advertising context [42], in a service encounter [9, 10] and finally on a website [20, 22]. Important limitations are nevertheless highlighted in a mere environmental approach [38, 39]. These show the need to take into account the important role of cognition, even within an environmental approach. First, there is a need to take into account the perceived congruency of the studied ambiance factors with its environment [40, in traditional environments; 20, for an online context]. Second the attitudinal reactions need not be totally discarded when environmental factors are studied, especially when, as here, attitudes are viewed as encapsulating not only cognitive but also affective components. In spite of these qualifications, a crucial contribution of such an environmental approach is the idea of approach-avoidance. This expresses the capacity of a design/ambiance factor to make people escape from, or affiliate towards an environment. [9, 10] synthesizes it as the will to stay, explore, return and affiliate. Such concept is related to the relationship building process, and perfectly matches the concept of stickiness used here. If this concept is transferred into man-machine interaction, it corresponds to what research in HCI has called “interface acceptability.” That is, an interface which generates behavioral stickiness and also positive, attitudinal on the one hand, and intentional on the other hand, reactions. Finally the concept of approach-avoidance reactions derived from environmental psychology, here conceptualized as the interface stickiness and operationalised by two measures of actual stickiness, and two measures of intentional stickiness, is actually an operationalisation of what is often called interface acceptability.

Last, an interface user actually socializes with the interface [45, 46] and this is particularly true when it displays an EVA. Agents have the aim of generating behaviors of actual stickiness [25, 26] and such positive effects are observed in a number of situations [11]. Positive effects are also observed, either on attitudes and intentions [15], in term of behaviors [8] or of persuasion [7].

3 Modeling and Testing Hypothesis

3.1 Model and Hypothesis

The model generally posits positive effects of an EVA on affective reactions, on attitudes, and through them on online stickiness. An EVA should have positive effects on affective reactions first, and through them on stickiness. Positive effects of EVAs are also posited on attitudinal reactions, which should partly explain effects on behavioral and intentional stickiness. That is, effects on stickiness may appear either in a stimulus-response approach (Hypothesis H1 to H4), through an attitudinal route (H5 to H18) or an affective route (H19 to H22), of influence (Figure 1):

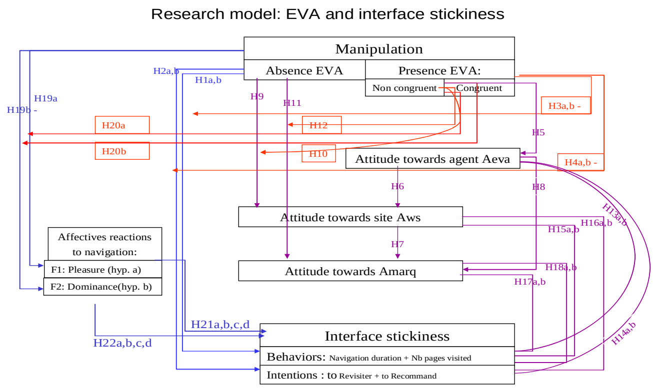


Fig. 1. Research model

3.2 Description of the Experiment

A laboratory experiment has been conducted in 2004 with 300 randomly recruited subjects, in the France Telecom R&D Center, HCI Division, near Paris (France). Subjects were exposed to one of two real brand sites and to 4 virtual agents (or no agent) created ad-hoc, 2 agents being tested on each sites . The log files, which allow behavioral stickiness measurement though the navigation duration and number of pages visited, could be collected for one site only. The two tested sites were: PRIMOLEA, presenting French high quality olive oil, and TRASER, promoting diving watches.

3.3 Results of the Research and Discussion

The constructs measurement and validity

The used measurement scales range from 5 to 21 items per construct: for Attitude towards the Agent AEVA, Attitude toward the site As, Attitude towards the brand Ab, Emotional reactions, agent Congruency with the website, and website Stickiness. All constructs are related to a relevant literature in psychology, man-man and man-machine interaction [23]. The measurement scales used here were developed by standard methods. All attitudes and congruency are unidimensional. Affective reactions are bidimensional: the dimension of Stimulation disappears in the factor analysis but Pleasure and Dominance – which has been most often eliminated in previous research – result in two, weakly correlated, dimensions. The nomological, convergent and discriminant validity of all here above proposed constructs are confirmed. Last, behavioral stickiness components are assessed through data extracted from the log files collected during the navigation.

Simple effects of the agent on stickiness

Effects on behavioral stickiness (only one site): the presence of an agent has a very significant effect on navigation duration (HIa***) and a significant effect on the number of pages (H1b*); but the congruence of the agent with the interface (here the site and the brand) has no significant effects on navigation duration (H3a NS, 5%<p)

nor on the number of pages (H3b NS¤, $5\%<p<10\%$). Effects on intentional stickiness (two sites): the presence of an agent has no significant effect on both components of intentional stickiness, the intention to revisit (H2a NS) and the intention to recommend (H2b NS). The effects of the congruence of the agent are also non-significant on both components of intentional stickiness in the aggregated navigations, but non-homogeneous across the sites: effects are not significant and positive on PRIMOLEA site (H4 NS), negative and significant on TRASER site (non-H4*), and globally not significant and negative on the aggregated navigation on both sites (H4 NS). Results are similar if we divide intentional stickiness into its two components, intention to revisit and to recommend.

Effects observed into and through the attitudinal hierarchy of effects
For the attitudinal hierarchy of effects (H 5 to H18, in the right part of the model) the effects of presence/congruency of the agent on the three attitudes (H5, H9 to H12) are examined. In case of presence of an agent: the agent congruency has no significant effect on the attitude towards the agent (H5 NS) and the observed effect of the agent congruency is negative (even if non significant), which is not expected. That is, users show a more positive attitude towards a surprising, non-congruent agent. For the two other attitudes Aws (towards the site) and Ab (towards the brand), the effects of agent presence are positive and highly significant (effect of presence on Aws: H9**, effect on Ab: H11***) but non-homogeneous across the sites. Last, the effects of agent congruency are not significant; and again non-homogeneous across both sites (effect of congruency on Aws H9 NS, and on Ab: H12 NS). Globally the effects of the presence of an agent are then confirmed, but not those of congruency. However, looking more closely into hierarchy of effects of attitudes, among the attitudes and then from each attitude towards stickiness reveals more subtle effects. The effects of each of the attitudes on the following one are validated, as well as the effects of each attitude on the intentional and on the intentional stickiness. But the effects of each attitude on the behavioral components of stickiness are not validated. When hypothesis are validated (i.e. among attitudes and from attitudes on intentional stickiness) close levels of explained variance and of correlations across both sites are observed, and above all, with high to very high effects. The correlations here range from 0.40 to 0.80. Since all constructs have adequate discriminant validities, [2], such high correlations are not problematic. The effects are quite homogeneous, without being equal, across the two tested sites, as the Chow tests conducted on each step of the hierarchy demonstrate. That is, the site in itself seems to have little effect on the validation of such route of influence of the agents. Last, the hierarchy explains quite well the effects on attitudes and on intentional stickiness, but not on behavioral stickiness. We had observed effects in a simple approach, on behavioral stickiness and those do not pass through the attitudinal hierarchy.

Effects observed into and through the affective route of influence
Also tested is the possible affective route of influence (H19 to H22) visible in the left part of the model. Interestingly, the results are quite similar to those observed in the attitudinal hierarchy. The first step of the sub-model is not confirmed: the agent presence and congruency have no significant impact on affect (H19 NS, H20 NS). That is, the manipulation of agent presence/congruency had very surprisingly no significant effects on affective reactions. The second step of the route is partly

validated on intentional stickiness only. Pleasure has a very significant effect on intentional stickiness (H21c,d***) and Dominance has a significant effect on one component of intentional stickiness (H22c NS on the intention to revisit, H22d* on the intention to recommend). But both affective reactions have no significant effects on behavioral stickiness (H21a,b NS, H22a,b NS). Last, correlations between attitudes (AEVA, Aws, Abrand) and Pleasure and Dominance are significant (Table 1).

Table 1. Correlations between Affect and Attitudes

Correlations (Pearson) AFFECT/ATTITUDES, aggregated navigations	Sig. level, r_{st} :
Pleasure <+> A_{EVA}	(**) +0.46
Dominance <+> A_{EVA}	(*) +0.13
Pleasure <+> A_s	(**) +0.66
Dominance <+> A_s	(**) +0.28
Pleasure <+> A_{brand}	(**) +0.55
Dominance <+> A_{brand}	(**) +0.23

As a conclusion, simple effects of the agent presence/congruency are observed on behavioral stickiness in the simple-effect approach. But if the approach is refined and two possible routes of influence are studied instead, significant effects are observed only on intentional stickiness, via the attitudinal route and via the affective route. Such routes do not capture all possible sources of influence and the proposed model may not include enough factors.

4 Limits of the Research and Contributions, Research Avenues

In this paper a theoretical framework is proposed that might help explain the observed effects of the agent, via an attitudinal route and an affective route. Most prior research on the effects of embodied agents does not propose a theoretical framework, but instead asserts that the agent must be "natural" and/or the interface "acceptable," leaving the actual process of influence largely a black-box. The framework proposed here partly explains the observed effects of embodied agents. It relies on previous research in traditional and online-based communication in marketing and psychology, in human-to-human communication, and in man-agent interaction, and seeks to go beyond the black-box approach to suggest routes of influence. A number of hypotheses are proposed and some effects are not significant on the sites taken separately, some effects are even significant in a direction opposite to what was expected. Above all, the two routes of influence do not account for the effects on behavioral stickiness. They do however, account for the effects on intentional stickiness. While this is an encouraging start, it does mean that the framework and model, is as yet incomplete in some areas.

Designing an EVA encompasses many possibly influencing variables: voice (and its many cues), non-verbal language, colors, the agent size, its functionalities and level of intrusiveness, its gender, age, personality, et cetera. All of these are potentially have impact on stickiness. Equally the gender, age, internet expertise–a construct still to be refined–of the users could also have impact. And all these

variables could be taken independently from the site, or taken into account through their congruency with the site. Thus a large field of questions and research is still open. Last, one limitation is inherent to any experiment: a number of variables, such as the real navigation on the site and exposition to the agent, gender, age, have been controlled; but in spite of a large sample (about 392 navigations, among which 344 valid navigations) our ability to generalize these results is limited by the fact that we cannot know if other non-controlled variables have had an effect, such as for instance and without being exhaustive: proneness to innovation adoption, attitude towards internet in general, social and leisure uses of internet with avatars, etc. All these limitations of our research are research avenues which might help better modeling man-agent interaction, taking into account more characteristics of the agent and of the user herself.

References

1. Richard, B.: The role of psychophysiology in consumer research. In: Robertson, R., Kassarjian, H. (eds.) Handbook of Consumer Behavior, pp. 124–161. Prentice Hall, Englewood Cliffs (1991)
2. Bagozzi, R., Yi, Y.: Evaluating structural equation models with unobservable variables and measurement errors: A comment. Journal of Marketing Research 18, 375–381 (1981)
3. Bagozzi, R., Gopinath, M., Nyer, P.: The role of emotions in marketing. Journal of the Academy of Marketing Science 27, 184–206 (1999)
4. Rajeev, B.: The Role of Affect in Consumer Behavior, Robert, P., Wayne, H., William, W. (eds.) pp.17–52 (1986)
5. Rajeev, B., John, M., David, A.: Advertising Management. Prentice Hall, Englewood Cliffs (2002)
6. George, B., Michael, B.: Advertising and promotion: An integrated marketing communications perspective. McGraw Hill, New York (1998)
7. Bjorn, B., Judee, B., Carl, C., Joseph, B., Magnus, L.: The impact of Anthropomorphic interfaces on influence, understanding, and credibility, (Hawaii International Conference on Computer and Systems Sciences, Maui, HI, &) Journal of Consumer Research, p. 14 (2001)
8. Timothy, B.: Relational agents, Effective change through human-computer relationships, PhD Dissertation, p. 284. MIT, Medialab (2003)
9. Jo, B.M.: Evaluating service encounters: The effects of physical surroundings and employee responses. Journal of Marketing 54, 69–92 (1990)
10. Jo, B.M.: Servicescapes: The impact of physical surroundings on customers and employees. Journal of Marketing 58, 57–71 (1992)
11. Scott, B., Clifford, N., Kevin, H.: Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent, International Journal of Human-Computer Studies, 62(2), 161–178 (2005)
12. Bruner, G.C., Kumar, A.: Web commercials and advertising hierarchy-of-effects. Journal of Advertising Research 40(1/2), 35–42 (2000)
13. Bucklin, R.E., Sismeiro, C.: A model of web site browsing behavior estimated on clickstream data (working paper 2002, et). Journal of Marketing Research 40(3), 249–267 (2003)
14. Qimei, C., William, W.: Attitude toward the site, Journal of Advertising Research, 27–37 (October 1999)

15. Jyun, C.Y., Gordon, M., Franck, B.: The effects of Anthropomorphic Agents on advertising effectiveness and the mediating role of presence. Journal of Interactive Advertising 2(1), 14 Fall (2001)
16. Tim, C., Timothy, D., Alopi, L., David, M.: E-Business: Revolution, Evolution or Hype? California Management Review 44(1), 57–86 (2001)
17. Derbaix, C., Pham, M.: Pour un développement des mesures de l'affectif en marketing: synthèse des pré-requis. Recherches et Applications en Marketing 4, 71–87 (1989)
18. Christian, D., Michel, P.: Affective reactions to consumption situations: A pilot investigation. Journal of Economic Psychology 12(2), 325–365 (1991)
19. Derbaix, C.: The impact of affective reactions on attitudes toward the advertisement and the brand. Journal of Marketing Research 32(4), 470–479 (1995)
20. Diesbach-Lambert, P.B.: Ambiance factors and approach behaviors in the e-commerce: A theoretical framework for capturing the effects of music and colour, In: Edwards, S. (ed.) Proceedings 1st Experiential E-commerce Conference, Michigan St.Univ. p. 14 (2001)
21. Diesbach-Lambert, P.B.: Ambience factors, emotions and web-user behavior: A model integrating an affective and symbolical approach, In: Gallopel, K., Maille, V., Rieunier, S. (eds.) 1st Colloquium on Sensorial Marketing, CERAM, Nice, p. 19 (2002)
22. Diesbach, P.B., Jeandrain, A.C.: Online immersion and emotional reactions online: proposal of an integrative model, p. 20. Asian ACR, Seoul (2004)
23. Diesbach, P.B.: Virtual agents and brand website stickiness, PhD Dissertation, IAE Aix-en-Provence/ESSEC, p. 684 (October 2006)
24. Diesbach, P.B., Galan, J.P.: L'agent virtuel incarné dans la distribution en ligne: cadre théorique et revue de littérature, Lemoine, J.F. (ed.) Univ. of Nantes, p. 24, (September 2006)
25. Judith, D.: Mediated faces, in Cognitive Technology: Instruments of Mind, In: Beynon, M., Nehaniv, C.L., Dautenhahn, K. (eds.) Proceedings of the 4th International Conference, CI 2001, Warwick, UK, p. 18 (August 2001a)
26. Judith, D.: Being real: Questions of Tele-Identity. In: The Robot in the Garden, Ken Goldberg, pp. 296–311. MIT Press, Cambridge, MA (2001b)
27. de Figueiredo, J.: Finding sustainable profitability in electronic commerce. In: Brynjolfsson, E., Urban, G. (eds.) Strategies for E-business success, pp. 7–33, 300, MIT Press, Cambridge, MA (2001)
28. Andrea, G.: Evolution of retail categories - an explanation from consumer's point of view. European Advances in Consumer Research 2, 237–245 (1995)
29. Groeppel-Klein, A.: The influence of the dominance perceived at the point of sale on the price-assessment. European Advances in Consumer Research 3, 304–311 (1997)
30. Groeppel-Klein, A., Thelen, E., Zandretter, C.: The Impact of Shopping Motives on Store Assessment, vol. 4, pp. 63–76. European ACR, Provo, UT (1999)
31. Andrea, G.-K.: Findings of environmental psychology for differentiation strategies in international retailing, In: Yves, D. (ed.) International Marketing: New frontiers and new tendencies, pp. 489–504, (& EIASM Seminar, Brussels 2000) (1998)
32. Donna, H., Thomas, N., Patrali, C.: Commercial scenarios for the Web: Opportunities and challenges, Journal of Computer-Mediated Communication, Special Issue on Electronic Commerce, p. 20 (1995)jcmc.huji.ac.il.volI/issue3/hoffman.html
33. Donna, H., Thomas, N.: Marketing in Hypermedia computer-mediated environments: Conceptual foundations. Journal of Marketing 60, 50–68 (1996a)
34. Donna, H., Thomas, N.: A new marketing paradigm for electronic commerce, Working Paper & Special Issue on Electronic Commerce, The Information Society (1996b), http://www2000.ogsm.Vanderbilt.edu/novak/new.marketing.paradigm.html

35. Donna, H., Thomas, N., Alladi, V.: Has the internet become indispensable? Communications of the ACM 47(7), 37–42 (2004)
36. Joseph, J.: Stickiness: A Ludicrous Metric, Avril (2002) (site http://www.mediapost.com/Admire/AdClickThru.cfm?AutoNumID=4247)
37. David, J.: Internet Retailing, ADMAP, In: Proceedings of the 6th ESOMAR International Automotive Marketing Conference, Lausanne, pp. 42–44 (July 2000)
38. Kellaris, J., Cox, A.: The effects of background music in advertising: a reassessment. Journal of Consumer Research 16(2), 113–118 (1987)
39. Kellaris, J., Cox, A., Cox, D.: The effects of background music on ad processing: a contingency explanation. Journal of Marketing 57(4), 114–125 (1993)
40. James, K., Susan, P.M.: Shaping time perceptions with background music: The effect of congruity and arousal on estimates of ad durations. Psychology and Marketing 13(5), 501–515 (1996)
41. Jonathan, K., Yongme, M., Rosalind, P.: This computer responds to user frustration: Theory, design and results. Interacting with Computers 14, 119–140 (2002)
42. MacInnis, D.J., Park, C.W.: The Differential Role of Characteristics of Music on High and Low Involvement Consumers' Processing of Ads. Journal of Consumer Research 18, 161–173 (1991)
43. Mehrabian, A., Russell, J.: Environmental Psychology, p. 221. MIT Press, Cambridge, MA (1974)
44. Albert, M.: Characteristic individual reactions to preferred and unpreferred environments. Journal of Personality and Social Psychology 46, 717–731 (1978)
45. Morkes, J., Flister, E.: Intuitive software, R&D Magazine, 66–69 (June 1999)
46. John, M., Kernal Hadyn, K., Clifford, N.: Effects of humor in task-oriented human-computer interaction and computer-mediated communication. HCI 14, 395–435 (1999)
47. Notebaert, J.F.: La création d'une relation entre le distributeur et le consommateur: une application au commerce électronique, PhD thesis, p. 347 (Novembe 2005)
48. Jean-François, N.: La création d'un lien social via internet, in 5ème Journée nantaise de E-Marketing, Lemoine, J.-F. (ed.) Université de Nantes (September 2006)
49. Martin, O., Jeff, M.: Capturing the consumer: ensuring website stickiness, ADMAP World Advertising Research Center, pp. 21–24 (August 2000)
50. Whan, P.C., Mark, Y.S.: Consumer response to Television commercials. Journal of Consumer Research 23, 11–34 (1986)
51. Rosalind, P.: Affective computing, p. 292. MIT Press, Cambridge, MA (1998)
52. Rosalind, P., Jennifer, H.: Affective Wearables, MIT Medialab Technical Report n. 467; and in Personal Technologies, 1, pp. 231–240 (1997)
53. Rosalind, P., Jonathan, K.: Computers that recognize and respond to user emotion, MIT Medialab Tech Report n.538; and in Interacting with Computers (2001)
54. Patricia, S.: Influence de la musique sur les comportements des acheteurs en grandes surfaces de vente, PhD thesis, Université de Rennes I, p. 638 (1994)
55. Julie, S., Gordon II, B., Anand, K.: Webpage background and viewers attitude, Journal of Advertising Research, 29–34 (April 2000)
56. Yoram, W., Vijay, M., Robert, G.: Convergence Marketing, Strategies for reaching the new hybrid consumer, p. 336. Prentice Hall, Englewood Cliffs (2002)

Integrative Physiological Design: A Theoretical and Experimental Approach of Human Systems Integration

Didier Fass

Artem Augmented Human project
ICN Business School and LORIA INRIA Lorraine
Campus Scientifique - BP 239 - 54506
Vandoeuvre-lès-Nancy Cedex
Didier.Fass@loria.fr

Abstract. Human modeling in design consists of human system integration (HSI), human factors integrated with systems engineering. That involves augmenting human capabilities and improving human-in-the-loop systems global performance, robustness and safety by behavioral technologies. For such human-in-the-loop systems design, this paper proposes an integrative physiological approach based on Chauvet's mathematical theory of integrative physiology (MTIP). By applying MTIP principles as theoretical framework, the integrative physiological modeling is used to model HIS and experiment a gesture-based method for virtual environment (VE) design and human system integration assessment. To demonstrate the pertinence and practicability of the developed integrative approach, we apply it to a wearable interactive system made up of virtual environment technologies for gesture assistance. The design prototype was evaluated in weightlessness during parabolic flights and confirms the effectiveness of the integrative physiological modeling.

Keywords: human modeling design, human system integration, augmented human, virtual environment, gesture assistance, weightlessness.

1 Introduction

1.1 Human Modeling in Design

The major benefits of using human modeling in design include reducing the need for physical development; reducing design costs by enabling the design team to more rapidly prototype and test a design; avoiding costly design 'fixes' later in the program by considering human factors requirements early in the design process; and improving customer communications at every step of product development by using compelling models and simulations.

Thus, designing an artifact consists in organizing a coherent relation between structures and functions in a culture and context of usage [design=structure/function]. Modeling human consists in taking into account anatomical and physiological elements in the same model. It is to design functions by organizing a hierarchy of structural elements and their functions [human modeling=physiology (functions)

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 52–61, 2007.

/anatomy (structures)]. Such models should be used to create models of individuals rather than using aggregated summaries of isolated functional or anthropometric variables that are more difficult for designers to use. Therefore human modeling in design requires an integrative approach [1].

1.2 Augmented Human

Augmenting cognition and sensorimotor loops with automation and interactive artifacts enhances human capabilities and performance. It is extending both the anatomy of the body and the physiology of the human behavior. Designing augmented human beings by using virtual environment technologies means to integrate both artificial and structural elements and their structural interactions with the anatomy, and artificial multimodal functional interactions with the physiological functions. Thereby the question is how to couple and integrate in a coherent way, a biological system with a physical and artifactual system, the less or more immersive interactive artifact, in a behaviorally coherent way by organizational design.

For training or operational systems design this requires taking into account technical devices, artificial multimodal patterns of stimuli alltogether, and their integration into the dynamics of the human sensorimotor behavior. Thus augmenting human capabilities and enhancing human performance by using virtual environment technologies needs safe design principles for human system integration (HSI). To be safe and predictive, HSI models, interaction and integration concepts, methods and rules have to be well grounded. Just like physical laws and their theoretical principals ground the mechanics or materials sciences and the engineering rules (e.g. for airplanes design), HSI needs a theory of integration, a theoretical framework and it's general principals.

In this paper, according to our epistemological approach, we propose to use the mathematical theory of integrative physiology (MTIP) [3] [4] as a fundamental framework for human system integration and virtual environment design [7]. For illustrating this new paradigm we present an experimental protocol using graphical gesture to assess HSI and VE design carried out both in laboratory and in weightlessness [6]

2 Epistemology

Converging technologies for improving human performances, *augmented human*, needs a new epistemological and theoretical approach of the nature of knowledge and cognition considered as an integrated biological, anatomical, and physiological process, based on a hierarchical structural and functional organization. Current models for human - machine interaction or human-machine integration are based on symbolic or computational cognitive sciences and related disciplines. They even use experimental and clinical data; they are yet based on logical, linguistic and computational interpretative conceptual frameworks of human nature where postulate s or axioms replace predictive theory. It is essential for the robust modeling and the

design of future rules of engineering for HIS, to enhance human capabilities and performance. *Augmented human* design needs both an integrative theory that takes into account the specificity of the biological organization of living systems, according to the principles of physics, and a coherent way to organize and integrate structural and functional artificial elements. Consequently, virtual environments design for augmented human involves a shift from a metaphorical, and scenario based design, grounded on *metaphysical* models and rules of interaction and cognition, to a predictive science and engineering of interaction and integration. We propose to ground HSI and augmented human design on an integrative theory of human being and its principles.

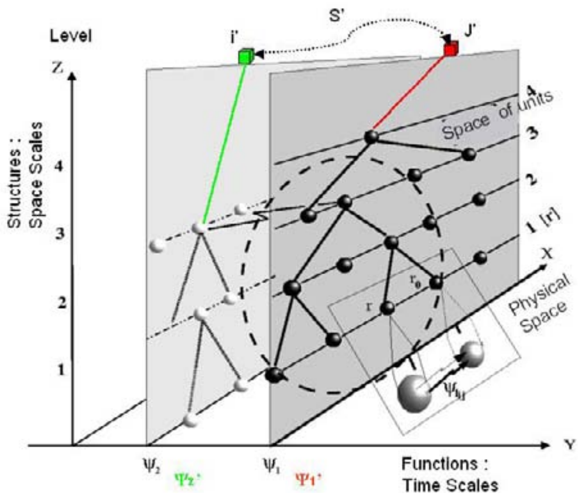


Fig. 1. 3D representation of the integrated augmented human design [7]

The human (Ω) is represented as the combination of the hierarchical structural (z) and functional (Y) organizations. The (x) axis corresponds to the ordinary physical or Cartesian space. Each physiological function ψ is represented in the $x\psi y$ plane by a set of structural units hierarchically organized according space scales. Two organizational levels are shown: ψ_1 and ψ_2. The different time scales are on the *y* axis, while space scales, which characterize the structure of the system, are on the *z* axis. The role of space and time clearly appears. Ψ_{1ij} is the non-local and non-symmetrical functional interaction. Units at the upper levels of the physiological system represent the whole or a part of sensorial and motor organs. Augmented human (Ω') design consists in creating an artificially extended sensorimotor loop by coupling two artifactual structural units *I'* and *J'*. Their integration into the physiological system is achieved by the functional interactions (i.e. sensorimotor) they generate. From sensors outputs to effectors inputs, the synchronized computerized process *S*' controls and adapts the integration of the functional interactions artificially created into the dynamics of the global and coherent system.

3 Mathematical Theory of Integrative Physiology (MTIP)

Mathematical theory of integrative physiology [2] [3] [4] examines the hierarchical organization of structures (i.e., the anatomy), and of functions (i.e. the physiology), of a living system and its behavior. It introduces the principles of a functional hierarchy based on structural organization within space scales, functional organization within time spaces of the structural unit that are the anatomical elements in the physical space. It copes with the problem of structural discontinuity by introducing functional interaction, for physiological function coupling, and structural interaction, for anatomical structural coupling. Unlike interaction in physics, at each level of organization, functional interactions are non-symmetrical, leading to directed graph, non local, leading to non local fields, and augment the system stability of a living system by coupling two structural elements. It is a new theoretical paradigm for knowledge-based systems and augmented human design.

MTIP is thus applicable to different space and time levels of integration in the physical space of the body and the natural or artificial behavioral environment; from molecular level to socio-technical level; from drug design to wearable robotics, and to life and safety critical systems design. Thus augmented human design (fig. 1) is grounding augmented human modeling on the Chauvet's MTIP.

This kind of model assumes the existence of functional interactions between the engineered and the physiological sensorimotor systems [7].

4 Experience

The gesture based method for virtual environment design and human system integration assessment is a behavioral tool inspired by Chauvet's theoretical framework, i.e.: (i) an integrated marker for the dynamical approach of augmented human design, and the search for interaction primitives and validation of organization principles; and (ii) an integrated marker for a dynamical organization of VE integrative design.

By designing a virtual environment, a human in-the-loop system consists in organizing the linkage of multimodal biological structures, sensorimotor elements at the hierarchical level of the living body, with the artificial interactive elements of the system, devices and patterns of stimulation. There exists a "transport" of functional interaction in the augmented space of both physiological and artifactual units, and thus a *function* may be viewed as the final result of a set of functional interactions that are hierarchically and functionally organized between the artificial and biological systems.

We present results of experiments performed by three subjects on earth and in hypergravity and weightlessness during parabolic flights (Flight campaign n°8 CNES-SPACEHAB), using a virtual and augmented reality system for gesture assistance. Using analysis of three-dimensional hand movements (drawing of ellipses), we compare the dynamical sensory-motor integration and motor performance (orientation, shape, figural and kinematics features) with or without the assistance of

virtual environments. Using this gesture-based method we evaluate physiological effects and integration of both change of gravity and artifactual environment on performance. We demonstrate how artificial visual information dynamically generated by a wearable virtual environment, may help gesture in the three-dimensional space parabolic flight, according to the MTIP principles

4.1 Integrative Physiological Design

Neurophysiology - Humans integrate multimodal sensorimotor stimuli in order to interact with their environment, be it natural or artificial (vision, vestibular stimulus, proprioception, hearing, touch, olfaction, taste..). When a subject is in a situation of weightlessness or of hypergravity, his sensorial system is submitted to an unusual pattern of stimuli. This dynamical pattern may largely influence the balance, the posture control, the spatial cognition and the spatial motor control of the subject [9]. If this coherence is absent, perceptual and motor disturbances appear, as well as illusions. These illusions are solutions built by the brain in response to the inconsistency between sensorial stimuli and internal processes. Therefore, the cognitive and sensory-motor abilities of the individual may be disturbed

Virtual environment as knowledge-based environment: Virtual reality and augmented reality technologies [12], because they are multimodal and aesthetic, are obviously the tools for the design and development of the assistance action and multimodal knowledge based artifactual environments. Knowledge is gathered from interactions and dynamics of the individual-environment complex and motivations. It is an evolutional, adaptable and integrative physiological process. It is fundamentally linked to emotions, mnesic process, perception and action. Then, designing an artifactual countermeasure system using virtual environment technology, a sensorimotor knowledge based environment, consists of making biological individual and artifactual physical system consistent. That needs an "eco-ethological" approach, both for the knowledge modeling, the interaction system design and the human system integration. Moreover, the coherence between artificial information and natural perceptual input is essential for the perception of space and the action within, especially during gravitational changes.

MTIP claims to develop experimental techniques to organize and to assess the behavioral coherence of the virtual environment design for augmented human performance. Because gesture is a high-level integrated sensorimotor and cognitive physiological function, it appears to be a primary expression of this global behavior and a behavioral tool for HSI and augmented human design[7].

4.2 Experiments

To test our assisting gesture prototype and highlight the dynamical principles of hierarchical organization of human systems integration and virtual environment design for assisting gesture in weightlessness, we set up a protocol according to a

complex and incremental design. The experiments were performed during the 8th CNES Novespace parabolic flight campaign.

*Devices***:** Head mounted display I-Glasses® immersive or see-trough, Frastrack Pohlemus® electromagnetic motion tracking system, workstation with a specific software design for managing and generating the visual virtual environment in real-time.

Protocol: Our protocol is based on graphical gesture analysis, more specifically of the drawing of ellipses within 3D-spaces. It's inspired by neurophysiology of movement [11] [9] [10]. Three right-handed trained volunteers were asked to draw ellipses (major axis 30 cm and minor axis 15cm) in two orientations of the three anatomical reference planes: vertical sagittal (VS) and transversal horizontal (TH). These drawing of ellipses were performed continuously and recorded during both the 1.8g ascent and the 0g parabola itself, feet in foot-strap (F) or in free-floating (FF), in two main situations: free gesture and assisted gesture wearing a visual virtual environment. Visual virtual environment was generated in immersion (RV) or in augmented reality (RA) (fig.2 and fig.3) .

*Data analysis***:** sixteen gesture-related variables are calculated from data produced during the parabola and recorded from the sensor worn on the tip of the index finger of the working hand: kinematics (Number of ellipses), Average velocity , Covariation Vt/Rt, Amplitude), position (Global position, Position / x axis, Position / y axis, Position / z axis), orientation (Global orientation, Orientation / sagittal plane, Orientation / frontal plane, Orientation / horizontal plane) and shape (Mean area, Eccentricity, Major axis variation, Minor axis variation).

Statistical analysis: We use a method of multidimensional statistical analysis. Principal component analysis and hierarchical classification are calculated with SPAD 4.0® to show the differential effects of hypergravity and microgravity on graphical gesture for each subject wearing or not the system. A second goal of this exploratory statistics is to assess the design of our prototype and the dynamics of the human virtual environment integration in weightlessness and on earth.

4.3 Results

Hypergravity (1.8g): Exploratory statistics show the effects of a prototype for assistance to gesture. Without assistance, the drawings of ellipses vary in an important way , both in shape (mean area, major and minor axis) and in position (global position and on the y axis). Orientation of gesture also has an influence. Global position and eccentricity present a larger variation in the TH orientation for the three operators. Inter-individual differences are a little more important in the VS orientation. The use of the gesture assistance improves the drawing of ellipses for both orientations with a difference between VR and RA, nevertheless. Graphical gesture is more accurate using immersive environment than augmented reality. We observe few inter-individuals variations and no orientation influence with virtual reality.

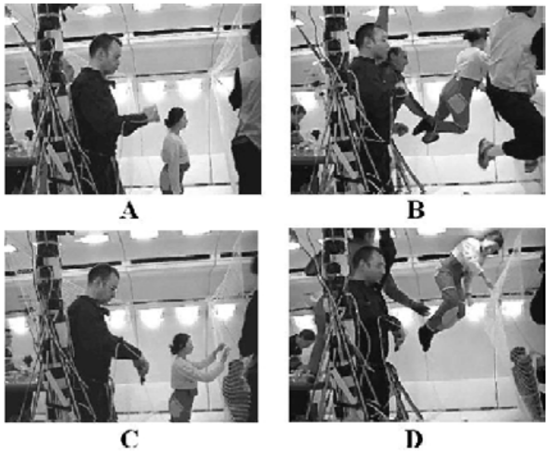


Fig. 2. Drawing of SV (A,B) and HT (C, B) ellipses without assistance in hypergravity (1,8g – A, C) and microgravity (0g – B,D)

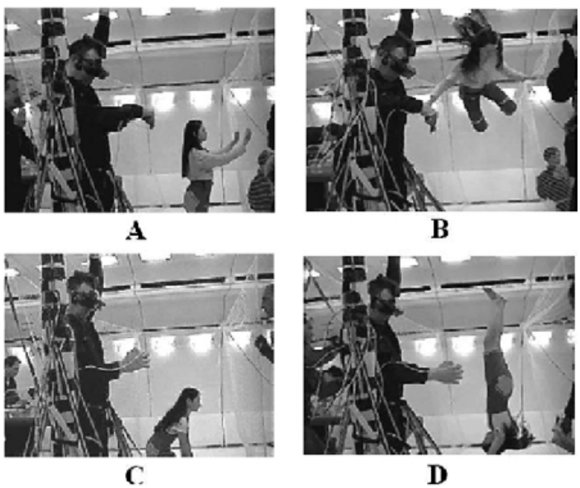


Fig. 3. Drawing of SV (A,B) and HT (C, B) ellipses with gesture assistance in hypergravity (1,8g – A, C) and microgravity (0g – B,D)

While with augmented realty assistance, see-trough helmet, gesture variations depend on both the individuals and the orientation. Differences are narrow for subject 1 and 2. Subject 3 presents broader variations of the global orientation in TH orientation regarding horizontal and frontal variations. In hypergravity, immersive assistance, virtual reality, is more efficient than augmented reality; moreover there are no differences as far as the foot-straps is concerned.

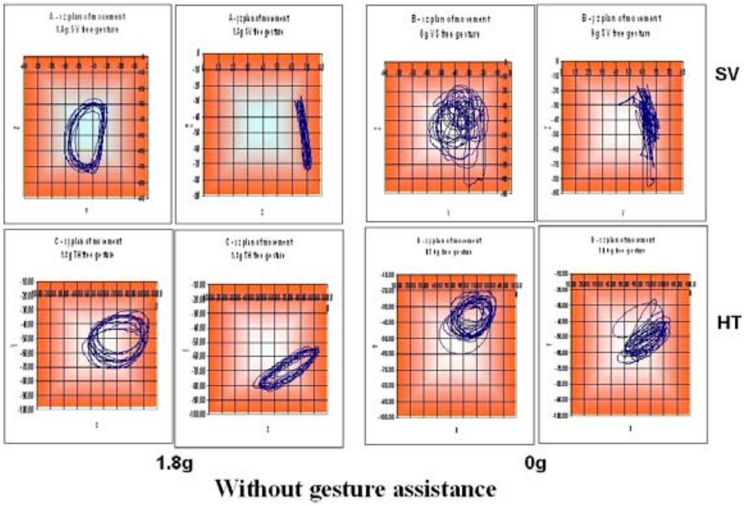


Fig. 4. Example of graphical gesture plotting of SV and HT ellipses drawn without assistance in hypergravity (1,8g) and microgravity (0g)

Microgravity (0g): Exploratory statistics prove the efficiency of virtual reality for assistance of graphical gesture in weightlessness. Immersive virtual environment improves position and orientation of drawn ellipses significatively and operator's behaviors are quite similar with feet in foot-straps. In free floating, there is also a homogeneous behavior in the VS orientation with VR assistance, but there are more inter-individual differences in TH orientation. In micro gravity our RA prototype brings less improvement than VR. RA main efficiency is on spatial orientation of gesture, especially on VS. For kinematics, position and shape variables, RA efficiency depends on both gesture orientation and individuals, with no influence of feet strapping.

In microgravity, the main inter-individuals differences depend on spatial orientation wearing RA assisting system or without assistance. For these two situations, subject1 presents fewer differences than the two others. His behavior is homogenous for both orientations and feet strapped or not. For TH orientation s1 presents a quite similar behavior with both VR and RA. With RA graphical gesture is close to s2 and s3 in VS orientation of action. Subjects s2 and s3 wearing RA systems are very sensitive to free floating. Their behavior is similarly dependant on gesture orientation with RA or without assistance.

Without assistance (Fig.4): Drawing of ellipses in hypergravity presents most changes of shape and position according to orientations and in VS plan of movement there are more inter-individual variations. Microgravity is mostly effecting orientation and position.

With assistance (Fig.5): Virtual environment, whether immersive virtual reality or augmented reality, improve graphical gesture in 3D space during parabolic flights. In hypergravity their effects are similar. In microgravity, VR is more efficient than RA especially concerning improvement of spatial orientation of gesture and position. Furthermore VR reduces inter-individual differences for VS and TH orientation with feet strapped and only for SV orientation in free floating.

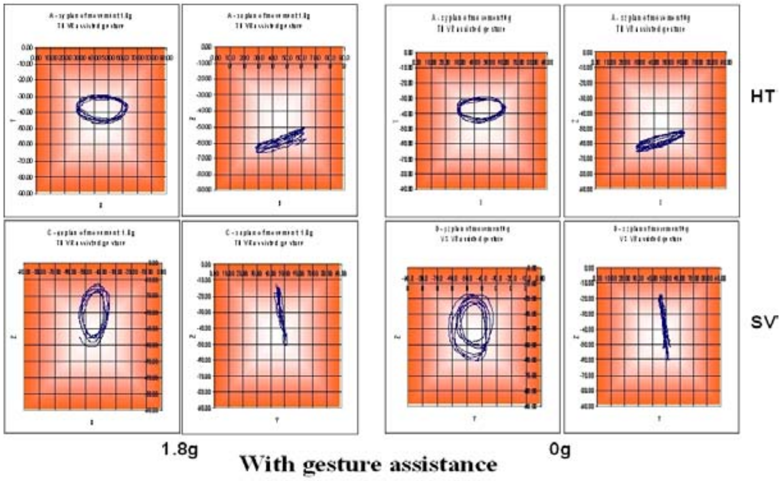


Fig. 5. Example of graphical gesture plotting of SV and HT ellipses drawn with assistance in hypergravity (1,8g) and microgravity (0g)

These experiments show (Fig.4 and Fig.5) the pertinence and practicability of the developed integrative approach of human modeling design for augmented human design and human system integration by showing improved motor skills and gesture performance.

5 Conclusions and Applications

Human modeling in design for the purpose of building human-in-the-loop systems can be a complex and dynamic endeavour. Augmented human design, i.e. assisting gesture technologies, needs an integrative approach that takes into consideration the specificity of biological organization of living systems, according to the principles of physics, and a coherent way to organize and integrate structural and functional artificial elements. Our experiments demonstrate the full potential of virtual environments for gesture assistance in weightlessness and in hypergravity. They assess the pertinence and the reliability of human systems integration modeling based on the MTIP principles.

Therefore, integrative physiological design is a framework for future developments of both augmented and cooperative human and environment. As virtual environments or wearable technologies, integrative artifacts will found the next assisting and countermeasures tools and smart environments for human space activities. Integrative physiological design will be necessary to model forthcoming architecture of safety critical systems or to develop applications to human on earth in the likes of surgeons, service engineers , physically challenged or elderly people.

Acknowledgements. We thank Pr G.A. Chauvet for his helpful comments on my research. Supported by CUGN, Région Lorraine, INRIA Lorraine.

References

1. Booher, H.R. (ed.): Handbook of human systems integration. Wiley, Hoboken (2003)
2. Chauvet, G.A.: Hierarchical functional organization of formal biological systems: a dynamical approach. I. to III. Philosophical Transaction of the Royal Society London B 339, 425–481 (1993)
3. Chauvet, G.A.: Theoretical Systems in Biology: Hierarchical and Functional Integration. In: Organisation and Regulation, vol. III, Pergamon, Oxford (1996)
4. Chauvet, G.A.: A new paradigm for theory in Integrative Biology: The Principle of Auto-Associative Stabilization. Biochemical networks and the Selection of Neuronal Groups. Journal of Integrative Neuroscience 5(3), 381–416 (2006)
5. Fass, D.: Graphical gesture analysis: behavioral tool for virtual environment design, Measuring behavior 2005. In: 5th International conference on Methods and techniques in behavioral research, 30 august to 2 september 2005, Noldus Information Technology, Wageningen, The Netherlands pp. 233–236 (2005)
6. Fass, D.: Virtual environment a behavioral and countermeasure tool for assisted gesture in weightlessness: experiments during parabolic flight. In: Proceedings 9th European Symposium on Life Sciences Research in Space and the 26th Annual International Gravitational Physiology Meeting, Cologne, Joint life conference Life in space for Life on Earth, Cologne, Germany, p. 173 (26 June - 1 July, 2005)
7. Fass, D.: Rationale for a model of human systems integration: The need of a theoretical framework. Journal of Integrative Neuroscience 5(3), 333–354 (2006)
8. Gurfinkel, V.S., Lestienne, F., Levik, Y.U.S., Popov, K.E., et Lefort, L.: Egocentric references and human spatial orientation in microgravity : II body-centred coordinates in the task of drawing ellipses with prescribed orientation. Experimental Brain Research 95, 343–348 (1993)
9. Ghafouri, M.: Représentation sensori-motrices du geste graphique dans l'espace tridimensionnel. Phd thesis, Université Henri Poincaré, Nancy1 (1997)
10. Ghafouri, M., Lestienne, F.: Contribution of reference frames for movement planning in peripersonal space representation. Exp. Brain. Res. 169(1), 24–36 (2006)
11. Laquaniti, F., Terzuolo, C., Viviani, P.: The law relating kinematic and figural aspects of drawing movements. Acta Psychologica 54, 115–130 (1983)
12. Stanney, K.: Realizing the full potential of virtual reality: human factors issues that could stand in the way, In: 1995 IEEE Virtual reality annual international symposium, Research triangle park, North Carolina pp. 28–34 (March 11-15, 1995)

Chinese Color Preference in Software Design

Yan Ge[1,2], Ronggang Zhou[1], Xi Liu[1], and Kan Zhang[1]

[1] State Key Lab of Brain and Cognitive Science, Institute of Psychology,
Chinese Academic of Sciences,
Beijing, 100101
[2] Graduate University of Chinese Academy of Sciences,
Beijing, China, 100049
{gey,zhangk}@psych.ac.cn

Abstract. Three experiments were designed to investigate the color preference of User Interface of Chinese youth. Background color and foreground color, as well as their combinations were examined. The results showed that: 1, Blue, purple, gray-blue and cyan were the more popular background color. 2, the foreground color preference was influenced by the background color, but white, yellow series and green series were all popular foreground colors for the background color of blue, purple and gray-blue. The discrimination of participants for graphic was better than that for characters. Some mechanism and implications were discussed.

Keywords: Color preference; User interface; Background color; Foreground color.

1 Introduction

Psychologists have been taking great interesting on color preference for a long time, there were some studies on color preferences using printed cards in the past. Zhou & Zeng (1986) [1], Wang (1997) [2] investigated color preference of Chinese students in different years. They found there was some difference in different years, but something was stable, e.g. white, green and blue were popular colors for Chinese students all the time. Li (1990) [3] found color preferences were different for different objects; gender and age also influenced preference. Elli (2001) [4] found significant gender difference in color preferences of North American college students. Today, the computer becomes more and more popular in the society, the study on color of computer user interface also attracts the attention of psychologists, but most of the studies mainly focused on the visual performance (Zhu & Cao [5], 1994; Cao & Zhu [6], 1995; Kong et al. [7], 1999). No research was focused on color preference itself on computer interface of Chinese people systemically.

In this study, we investigated the color preference of User Interface of Chinese youth. We intended to use a better way to study this issue. First, 7~10 colors would be chose from the basic colors pool, using as materials of background color study. Second, background color preference was explored in details using forced choice and subjective evaluation. Third, combining foreground color, we investigated color

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 62–68, 2007.

combination preference. Finally, visual performance of color combination was also tested.

2 Experiment 1

2.1 Purpose

This experiment was designed to choose materials for experiment 2. Participants were asked to choose 10 favorite colors as background color of user interface.

2.2 Method

Participants. 20 students from Beijing Normal University (N=10 male, N=10 female), aged 20 to 23. All participants had normal color vision and normal or correct to normal visual acuity.

Materials. 48 basic colors from the drawing program of windows 98, each color was showed in 20×13mm rectangle, were arranged in a 8×6 matrix. All of them were presented on a screen. The background color was light gray.

Apparatus. The material was presented on a 17-inch color CRT monitor, 1024×768 pixels，saturation 100%，brightness 50%，red 55%，green 55%，blue 55%。

Procedure. We showed the 48 basic colors on the computer screen in one time, the participants were asked to choose 10 favorite colors from them as the background color of computer user interface and to order these colors by their preference.

2.3 Data Analyze and Results

42 colors of 48 basic colors were chose as favorite colors in this part. Each color was scored based on their order. For example, if one person chose blue as first color, it was scored 10; if one person chose blue as second color, it was scored 9; the rest may be deduced by analogy. The final score of each color was the sum of every participant's results. The first 10 colors are blue, dark purple, dark blue, gray, white,gray blue, purple, gray purple, cyan and rose. Their scores are 87, 82, 72, 62, 61, 60, 55, 54, 49 and 39 respectively.

A Chi-square test revealed there were significant difference among these 10 colors, χ^2=31.254，df=9，P<.001. Because the purpose of this study was to explore color preference of colorful user interface, gray and white were exclude in next experiment. Finally, blue, dark purple, dark blue, gray blue, purple, gray purple, cyan and rose color were chosen to be the material of experiment 2.

3 Experiment 2

3.1 Purpose

This experiment, applying forced choice and subjective evaluation, was designed to investigate the background color preference based on the results of experiment 1.

3.2 Method

Participants. 44 students from China Agricultural University and Institute of psychology (N=23 male, N=21 female), aged 19 to 34, joined in this experiment. All participants had normal color vision and normal or correct to normal visual acuity.

Materials. 8 colors chose in experiment 1 were used to be the materials.

Procedure. Experiment 2 was composed of two parts: Part one: forced choice. In this part, two colors of 8 colors were presented on a screen each time; each color took up half of the screen. Participants were asked to choose their favorite background color in the two colors. If they chose left color, they should press “1” on the number keypad; if they chose right color, they should press “2” on the number keypad. In order to exclude location effect, each color compared with other 7 colors both on left and right. The choice time of each color was recorded. There were four times practices before the formal experiment. Part two: subjective evaluation. Participants were asked to rate each color in enjoyable, comfortable and suitable as background color of a computer interface. 7-point scale were used in this part (1=most enjoyable/comfortable/suitable, 7= most non- enjoyable/comfortable/suitable). One color presented randomly on the whole screen each time, three scales were showed on right of the screen. Participants were asked to use mouse to choose corresponding score for each color. After evaluated three aspects, participants can press the “NEXT” button at the bottom of the screen to go to evaluate next color.

3.3 Results

Forced Choice. Each participant’s choice times were transformed to Z score, showed in table 1. Repeated measures of General Linear Model were used to analyze the difference among different colors, $F(7, 301) = 10.223$, $P<0.001$.

Table 1. Descriptive results of forced choice

color	blue	purple	cyan	gray blue	dark purple	dark blue	gray purple	rose
Z score	0.52	0.36	0.30	0.25	-0.18	-0.38	-0.47	-0.48
Z’	2.35	2.19	2.13	2.08	1.65	1.45	1.36	1.35
SD	0.49	0.38	0.90	0.59	0.67	1.13	0.97	0.95

Pairwise comparisons were used to compare the Z score of each color. The results showed there was significant differece between blue and last 5colors. It was obvious that blue was the most favorite background color for participants. Purple, cyan and gray blue were also different with last 4 colors. It showed that they were also the favorite color for participants.

Subjective evaluation. The result of subjective evaluation in three aspects was showed in fig.1.

In the enjoyable aspect, the order of background color preference was blue, gray blue, purple, cyan, rose, dark purple, dark blue and gray purple. The results of repeated measures of General Linear Model showed there was significant difference among 8 colors, $F(7,301) = 8.895$, $P<0.001$. Pairwise comparisons results showed there was significant difference between blue and other 7 colors, blue was the most enjoyable background color.

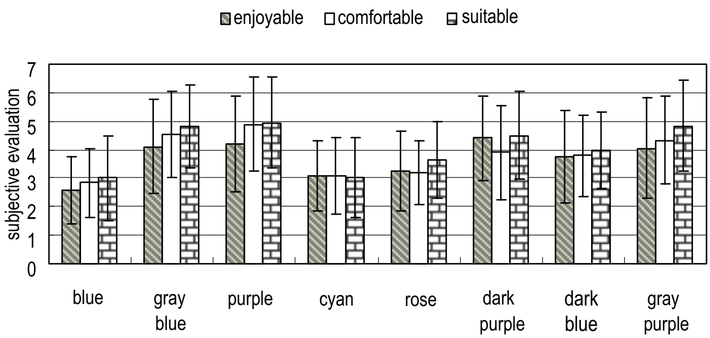


Fig. 1. Subjective evaluation of background colors

In the comfortable aspect, the order of background color preference was blue, gray blue, purple, cyan, gray purple, rose, dark purple and dark blue. The results of repeated measures of General Linear Model showed there was significant difference among 8 colors, $F(7,301) = 13.960$, $P<0.001$. Pairwise comparisons results showed there was significant difference between blue,gray blue,purple and other 5 colors. Blue, gray blue and purple were more comfortable background color.

In the suitable aspect, the order of background color preference was blue, gray blue, purple, cyan, gray purple, dark purple, rose and dark blue. The results of repeated measures of General Linear Model showed there was significant difference among 8 colors, $F(7,301) = 16.038$, $P<0.001$. Pairwise comparisons results showed there was significant difference between blue, gray blue, purple, cyan and other 4 colors. Blue, gray blue, purple and cyan were all more suitable as background colors.

The results in three aspects were similar, Spearman correlation coefficient was used to test the correlation of three aspects, there were significant correlation between each two aspects, $r_{enjoyable\times suitable} = 0.932$, $r_{comfortable\times suitable} = 0.959$, $r_{enjoyable\times comfortable} = 0.875$. It prompted they could impact each other, so they all need to be considered in the interface design.

Both forced choice and subjective evaluation results showed blue, gray blue, purple and cyan were more favorite to use as background color, so they were used as background color in the next experiment, to explore color combination preference.

4 Experiment 3

4.1 Purpose

Subjective evaluation was used to investigate the color combination preference.

4.2 Method

Participants. 43 students from China Agricultural University (N=23 male, N=20 female), aged 19 to 24, joined in this experiment. All participants had normal color vision and normal or correct to normal visual acuity.

Materials and procedure. Blue, gray blue, purple and cyan were used as background colors, other basic colors were used as foreground color. The procedure was similar as subjective evaluation part of experiment 2. One foreground color was showed in 20×13mm rectangle on one background color each time, participants were asked to rate if this combination was suitable as computer user interface on a 7-point scale (1=most suitable; 7= very not suitable). They need to evaluate 184 times at all.

4.3 Results

Table 2 showed first 10 color combinations.

Table 2. Descriptive results of first 10 color combinations

Foreground Background	Blue/ white	Blue/ light yellow	Blue/ cyan	Blue/ grass green	Gray blue/ white	Blue/ light green	Purple/ white	Purple/ light yellow	Blue/ green	Blue/ yellow
Mean	2.95	3.26	3.28	3.37	3.37	3.4	3.47	3.51	3.53	3.53
SD	1.59	1.57	1.71	1.9	1.96	1.68	1.86	1.67	1.91	1.72

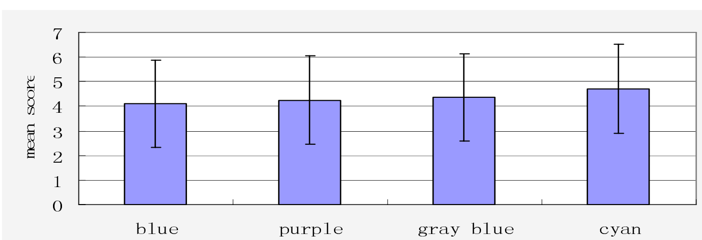


Fig. 2. Subjective evaluation among different background color

It was obvious that blue was the favorite background color, purple was also popular. Cyan was not chosen into first 10 combinations. Repeated measures of General Linear Model were used to analyze the difference among four background

colors, descriptive results showed in fig.2, $F(3, 126)=6.971$, $P<.01$, the combinations in which blue was used as the background color were the favorite combinations for the participants. It prompted that the color combination preference was impacted by background color.

For the foreground colors, white series, yellow series and green series were all popular foreground colors when the background color was blue, purple or gray-blue. But these results were just observed from descriptive results, more research was needed in the future.

5 Discussions

Results of background color preferences experiment showed, blue, purple, cyan and gray blue were chosen as participants' favorite background colors for computer user interface. Although the order of these 4 colors in two methods was not all the same, there were significant difference between them and other 4 colors. Blue was the most popular background color, results of forced choice and subjective evaluation were both proved this. This may be related with the characteristic of blue, some research prompted blue can inspire positive emotion, e.g. comfort, stillness et al. (Wu & Wang[8], 1986; Huang et al.[9], 1991). Another reason was participants' habit, the most popular computer user interface, which was used in China widely, was using blue as its main hue (e.g. Windows 98, 2000, XP). Three aspects of subjective evaluation revealed suitable background color was high related with its comfortable and enjoyable degree. If a color was suitable as background color of interface, participants would consider if it was comfortable and if he/she like it. There were some prompts for interface design. But Zhu and Cao (1994) found there was some discord in definition and comfort. How to apply these factors properly was a meaningful question.

Based on the results of color combination preference, foreground color preferences were influenced by background color, blue was still the most favorite background color. This was coincident with common computer interface (Microsoft Windowed series). The foreground color preference was influenced by the background color. White, yellow series and green series were preferred to be favorite foreground colors. White with blue, white with gray blue and white with purple were the best combination in these three background colors. These rules needed to be considered in interface design.

6 Conclusions

In a word, our works indicated that blue was the best background for young computer users, and purple, gray-blue and cyan were also good choices. Color combination preference was influenced by background color. White series, yellow series and green series were all popular foreground colors for the background color of blue, purple or gray-blue.

References

1. Zeng, H.W.: Color preference of Chinese students and its development [J]. Psychology science communication (in Chinese) 9(1), 16–22 (1986)
2. Wang, Y.R.: An experimental study of color preference of Chinese University students [J]. Psychology science (in Chinese) 20(6), 559–560 (1997)
3. Li, W.F.: Color preference of Chinese townsmen: the relationship between color preference and some factors [J]. Psychology science communication (in Chinese) 13(5), 13–19 (1990)
4. Ellis, L., Fricek, C.: Color preferences according to gender and sexual orientation [J]. Personality individual differences 31(8), 1375–1379 (2001)
5. Zhu, Z.X., Cao, L.R.: The effect of object-background color combination to visual efficiency of color CRT [J]. Transactions of Psychology (in Chinese) 26(2), 128–135 (1994)
6. Cao, L.R., Zhu, Z.X.: The effect of contrast to visual efficiency of color CRT [J]. Ergonomics (in Chinese) 1(1), 32–36 (1995)
7. Kong, Y., Ge, L.Z., Wang, Y.J.: A comparison study of salient efficiency of 4 colors on black and white background [J]. Ergonomics (in Chinese) 5(4), 12–14 (1999)
8. Wu, X.Z., Wang, X.Y.: An investigation of “color-emotion” in university students [J]. Psychology science communication (in Chinese) 9(1), 38–40 (1986)
9. Huang, X.T., Huang, W., Li, X.R.: A color-emotion study of Chinese people [J]. Psychology science (in Chinese) 14(6), 1–7 (1991)

The Effect of Animation Location and Timing on Visual Search Performance and Memory

Songmei Han

Dept. of Psychology, SUNY Oswego, Oswego, NY 13126 USA
han@oswego.edu

Abstract. The current study investigated the effects of animation location and timing on visual search speed and accuracy and their effects on memory about the animated strings. Visual search accuracy was measured using the sensitivity measurement d' in signal detection theory (SDT) model. Results showed that black-and-white animations had no significant effect on visual search and color animations slowed down the search significantly but had no significant effect on search accuracy. The size of the effect that an animation had on the search speed did not depend on its location or timing. Nor did the ability to recognize the animated string or the preference judgment about the animated string depend on its location and timing. Animated strings were rated more preferable than new strings even in the absence of explicit memory about the animated strings.

Keywords: visual search, visual attention, animation, online advertisement, signal detection theory.

1 Introduction

Online advertising is growing at a fast pace. The Internet advertising revenues for 2005 totaled over 12.5 billion and increased over 30% from 2004 and accounted for 5% of total U.S. advertising revenues in 2005 [5]. The growth continued in 2006 and the Internet advertising revenues reached a new record of 7.9 billion for the first 6 months of 2006, a 37% increase over the first half of 2005 [6]. Online advertising can take many forms: static banners, pop-up cookies, and animations. Animations are more effective in online advertising because they improve click-through rate by 25% [7, 8] and might also enhance user's memory about the advertisement [1].

1.1 Effects of Animation Location and Timing

Most studies on online advertisement have focused on the effectiveness of the online advertisement without looking at the effects of online advertisement on user's task performance at the website. Zhang [10, 11, 12] conducted several studies on the effects of animation on information seeking performance measured by a visual search task where participants searched for a target string among distracter strings. Zhang found that animations had negative effects on visual search performance and the size of the negative effect depended on animation location and onset timing. Animations at

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 69–78, 2007.

the left side deteriorated performance more than those at the right side. Animations that appeared in the middle or close to the end of visual search deteriorated performance more than those at the beginning. Zhang explained these findings with visual attention theories. Due to the limited nature of visual attention, the more attention an animation received from a user, the less amount of attention was available for the visual search task and the more damaging it was to the search task. Since the left side of printed media or computer screen tends to contain more important information in English speaking countries, users tend to pay more attention to animation on the left side and performed worse in the search task when animations appear on the left side. Since most animations on the Internet appear when the web page is loaded, users do not expect to see animations in the middle of a search. The surprise caused by the animations in the middle of a search might be more distracting to the visual search task.

Zhang's studies examined the effects of animation location and onset timing on visual search task without measuring the effectiveness of the advertisements. It is unclear whether the more distracting animations were remembered better. The current study will answer this question.

A user's memory about the animations can be measured explicitly with recall and recognition tests and implicitly with preference judgment. Past studies found that attention had significant effects on explicit memory but very little or no effects on implicit memory [4]. Previous research [9] found that frequent exposure increased the appeal of various initially unfamiliar stimuli even when subjects were not aware of the difference in exposure frequency. User preference judgment can change even without the user paying attention to the advertisement. The amount of attention an advertisement receives should have no effect on preference judgment. If animation location and timing affect search performance by changing the amount of the attention devoted to the animation, they might affect recall and recognition of the animations but have no effect on preference judgment about the animated strings.

1.2 Visual Search Modeled with Signal Detection Theory (SDT)

Psychophysicists often use signal detection theory to model visual search performance because the model takes into consideration the external noise in the visual field and the internal noise in the perceptual system and neural system [2]. Signal detection theory allows psychologists to measure a participant's sensitivity in a visual search task without the contamination of his or her decision bias. An example can easily illustrate the distinction between an individual's relatively stable sensitivity and the constantly changing decision bias adapted by the person. Imagine you are the operator of the security check machine at an airport and has to decide whether to stop each bag for manual inspection. Dangerous items are the signals and similar looking harmless items are noises. Your ability to distinguish between these items is relatively stable but your criteria in stopping a bag can change instantly. When warned for a terror threat, you will be extremely cautious and will stop any bag containing items that look remotely like a dangerous item. By doing so, you not only will stop more bags containing dangerous items but also will stop more harmless bags. Stopping a

dangerous bag is called a hit and stopping a harmless bag is called a false alarm. Two measurements can be computed from a set of hit rate and false alarm rate to measure the operator's sensitivity and decision criteria separately. The d' measures the sensitivity and the β measures the decision criteria.

$$d' = z\,(\text{hit rate}) - z\,(\text{false alarm rate}) \tag{1}$$

The z-score is computed from a Gaussian distribution. In a visual search task that required participants to click targets among distractors, participants can click every item to achieve a hit rate of 100%. By doing so, they will also have a false alarm rate of 100%. The sensitivity (d') in this case will be zero.

Zhang calculated search accuracy according to the following formula [12]:

$$\begin{aligned} CA = \text{NumberOfClickedTargets} * \text{NumberOfClickedTargets} \,/ \\ (\text{NumberOfTargets} + \text{NumberOfWrongClicks}) \end{aligned} \tag{2}$$

Although the formula aims to correct false alarm, it is not a standardized measurement. The d' is a standardized measurement. The current study measures search accuracy with d'.

1.3 Speed and Accuracy Tradeoff in Visual Search

In a visual search task, participants can trade speed with accuracy by choosing to respond fast with lower accuracy or to make an accurate response with long response time. A complete measure of performance should look at both speed and accuracy. Zhang measured performance with the following formula [12], where CA referred to Click Accuracy:

$$P = 10000 * CA \,/\, \text{TimeOnTaskpage} \tag{3}$$

Although this measure is adjusted for the speed-accuracy trade-off in the task, it cannot tell us whether animation affects search time and accuracy differently. Past studies indicate that attention demanding tasks can slow down the information processing in visual search without reducing the asymptotic accuracy level [3]. The current study examines search speed and accuracy separately to test whether animation has different effects on different aspects of the search task.

1.4 Hypothesis

It is predicted that animations will slow down the search process but will not affect the search accuracy. It is predicted that the size of the damaging effect will depend on animation location and timing and the more damaging animations will be remembered better. Animation location and timing will have no effects on the preference judgment of the animated strings. Since color can increase the salience of an object, it is predicted that animations in color will be more distracting than animations in black and white and will be remembered better than animations in black and white.

Two experiments were conducted to test these hypotheses. Non-sense strings were used to control the familiarity effects on search and memory performance.

2 Experiment 1

2.1 Methods

Design. A 2 (location) x 4 (timing) factorial within-subject design was used. Animation appeared either on the left or the right side of the search screen. All animations lasted 10 seconds with four possible onset timing: when the search screen appeared, when participants clicked on any string in the second half of the screen, when participants clicked on any string in the last quarter of the screen, or on-off-on where animation started when the search screen appeared and stopped after 5 seconds and reappeared when the participants reached the last quarter of the screen and lasted for 5 seconds. A control condition with no animation was added to test whether animation had any effects on performance. Each participant performed search tasks under all nine conditions.

Participants. Forty-two students enrolled in psychology courses at a university in the United States participated in the experiment for course credits.

Materials. Each search screen consisted of 8 rows with 15 strings each in each row. Strings were generated by a computer program with a random mechanism. A 2-letter string was generated by picking the first and the second letter independently. A 3-letter string was generated by picking the first and the last letter from the consonants and the middle letter from the vowels. A 4-letter string was generated by adding a consonant to a 3-letter string. A target was always a 4-letter string. The same target appeared three times in each row with a total of 24 appearances. The positions of the target were chosen randomly with the constraint that no two targets would occupy adjacent locations in a row. Eighteen search screens with different targets were created independently.

Animation was created by increasing the font size of a 4-letter string from 24 to 72 points in eight steps (24, 28, 32, 36, 42, 48, 56, 64, 72) and cycling between two strings. The animation would increase the font size of the first string from 24 to 72 and then switch to the second string and increase the font size in a same manner and then switch back to the first string to start over. Eighteen animations were created. Animations appeared at the side of the screen near the top. Two sets of animation strings were created. During the experiment, half of the participants saw Set A as animation and half of the participants saw Set B as animation.

Both the search strings and the animated strings were black letters on a light gray background. The rows on the search screen were well-spaced to occupy the whole screen. The search was conducted in a desktop application instead of web browser.

Procedure. The experiment consisted of 18 trials with two trials for each experimental condition. Participants performed two practice trials at the beginning of the experiment. All participants searched the same screens. The order of the conditions, the search screen used for each condition, and the animation used for each condition was randomized for each participant.

Participants were shown the target string before each search and were instructed to click a button to start the search. Upon the click, a screen appeared with the target

string at the top center of the screen. Participants were instructed to click on target strings only and click all instances of the target string and to finish the search quickly. They were told that clicks on non-target strings and targets missed were counted as errors.

The time from the appearance of the screen to the click on the “Finish” button was recorded as search time. After clicking on the “Finish” button, the participant was given feedbacks on the number of total clicks, the number of correct clicks, and the number of incorrect clicks.

After finishing all conditions, participants were asked to write down any animated strings they remembered. They were then given a piece of paper containing 72 strings from Set A and Set B (32 strings animated during the experiment and 40 unseen strings) and were asked to circle those strings they saw in the animation during the experiment. At last, participants were asked to rate how much they liked each of the 72 strings on a 5-point scale (1 = dislike the most, 5 = like the most). The same set of strings was used in recognition and rating but the strings appeared in different orders.

2.2 Result

Data from nine participants showed extremely low accuracy with a hit rate of 33% or lower in one or more conditions. Their data were dropped from the data analysis. Data from 33 participants were used in the following analyses.

Effects of Animation on Visual Search. Paired-sample t-tests were conducted to compare the grand mean of all animation conditions with the control condition. Results showed no significant difference in p-measure, $t(32) = 0.937$, $p =0.356$, search time, $t(32) = -1.625$, $p = 0.114$, or sensitivity (d’), $t(32) = 0.357$, $p = 0.724$. No further analysis on the effects of animation location and timing was done.

Memory About Animated Strings. Participants performed very poorly on recall and recognition of the animated string with accuracy close to zero. No further analysis on the recall or recognition data was done.

The rating of the animated strings was compared with the rating of the new strings in a paired-sample t-test. The result indicated that animated strings were rated significantly higher than new strings, $t\ (32) = 2.884$, $p = 0.007$. The mean rating of the animated strings was 2.943 while the mean rating of the unseen strings was 2.832.

Results from a 2 x 4 repeated measured ANOVA on preference judgment showed no significant main effects of animation location, $F\ (1, 32) = 0.548$, $p = 0.465$, no significant main effect of animation timing, $F(3,96) = 1.243$, $p = 0.298$, and no significant interaction between animation location and timing, $F\ (3, 96) = 2.364$, $p = 0.076$.

2.3 Discussion

This experiment showed no detrimental effects on visual search performance from animations. This lack of animation effects might be because the animation did not attract substantial attention. The lack of attention to the animated strings also account for the extremely poor memory about the animated strings in recall and recognition.

Although the explicit memory about the animated strings was very poor, the implicit memory test (preference judgment) showed that mere exposure did increase the likeability of the animated strings. This result is consistent with previous findings that frequent exposure increased the appeal of unfamiliar stimuli [9].

The failure in replicating the detrimental effects of animation on visual search performance in Experiment 1 might be due to the fact that the search task was too difficult and the animation was too dull. Zhang found that the effects of animation decreased with the increase in task difficulty levels and that dull animation had less detrimental effects on visual search tasks than brightly colored animations [11]. In Experiment 2, color animation was used to make the experimental condition similar to the advertisement animation used on the Internet and to increase the likelihood that animation will have effects on the primary visual search task.

3 Experiment 2

3.1 Methods

Design. A 2 x 4 within-subject design with a control condition was used as in Experiment 1.

Participants. Forty-six undergraduate students enrolled in psychology courses at a university in the United States participated for course credits.

Materials. The materials in Experiment 1 were used in Experiment 2 with a few modifications. First, the targets were 3-letter strings instead of 4-letter strings. Shorter target strings were used in Experiment 2 to generate the optimal animation effect [11]. Second, the color instead of the size of an animated string was changed to produce animation. The size (72 points) of the string stayed the same but its color changed from blue, to green, to magenta, and then to orange. Third, the animation cycled among three strings instead of two. Fourth, there was no additional space between rows on the search screen.

Procedure. The procedure in Experiment 1 was used with one exception. A recognition test of the animated strings was given immediately after each search. After clicking on the “Finish” button, participants were shown a new screen with 6 strings and were asked to select strings animated during the search if there were animations. No feedback on the search or the recognition was given. A new trial started once the participant finished the recognition test. A blank screen with “New Search” button was used after a search with no animation.

At the end of the experiment, participants were given 108 strings from Set A and Set B (48 strings animated during the experiment and 60 unseen strings) and were asked to rate how much they liked each string on a 5-point scale (1 = dislike the most, 5 = like the most). The same set of strings was used in recognition and rating with different orders.

3.2 Results

Data from nine participants showed extremely low accuracy with a hit rate 33% or lower in one or more searches and one participant had difficult using the mouse during the experiment. Their data were dropped from data analysis. Data from 36 participants were analyzed.

Effects of Animation on Visual Search. The effects of animation were examined with paired-sample t-tests. Table 3 shows the results of the t-tests. Animation increased the time spent on the search and reduced performance significantly but had no significant effect on sensitivity (d').

Table 1. Effects of Color Animation on Visual Search

Measures	Animation	Baseline	t(df= 35)	*p*
p-measure	5616	6034	3.961	**0.000****
Search Time	38.089	35.893	-3.691	**0.001****
d'	6.44	6.47	0.174	0.863

Results from a 2 x 4 repeated-measure ANOVA on search time showed no significant main effect of animation location, F (1, 35) = 0.748, p = 0.393, no significant main effect of animation timing, F (2.374, 83.091) = 1.766, p = 0.171, and no significant interaction between animation location and timing, F (3, 105) = 1.237, p = 0.300.

Results from a 2 x 4 repeated-measure ANOVA on p-measure showed no significant main effect of animation location, F (1, 35) = 0.154, p = 0.697, no significant main effect of animation timing, F (3, 105) = 2.090, p = 0.106, and no significant interaction between animation location and timing, F (3, 105) = 1.395, p = 0.248.

Since animation had no significant effects on sensitivity (d'), no further analysis on sensitivity (d') was done.

Memory About Animated Strings. To analyze the recognition data, a hit rate and a false alarm rate were calculated for each participant in each animation condition. The recognition performance was measured by sensitivity (d'). Data from participants who had negative sensitivity (d') in a condition and participants who had zero sensitivity (d') in more than one conditions were dropped from analysis. Data from 28 participants were analyzed.

Results from a 2 x 4 repeated measure ANOVA on recognition showed no significant main effect of animation location, F (1, 27) = 0.154, p =0.698, no significant interaction effect between animation location and timing, F (3, 81) = 0.745, p = 0.529 and a significant main effect of animation timing, F (3, 81) = 5.522, p =0.002. Six pairwise comparisons among different animation timing were conducted with Bonferroni adjustment with a significance level of 0.05. Results showed that strings animated close to the end of the search were recognized significantly better those in on-off-on animations. The differences between other animation timings were not significant.

Animated strings were rated significantly higher than new strings, t (34) = 3.630, p = 0.001. The mean rating of the animated strings was 2.675 while the mean rating of the new strings was 2.543. Results from a 2 x 4 repeated measure ANOVA on preference judgment showed no significant main effect of animation location, F (1, 34) = 2.170, p = 0.150, no significant main effect of animation timing, F (2.452, 83.363) = 1.344, p = 0.313, and no significant interaction between animation location and timing, F (3, 102) =2.635, p = 0.054.

3.3 Discussion

The Effects of Animation on Visual Search Performance. Color animations slowed down visual search and reduce performance (p-measure) but had no effect on sensitivity (d'). The detrimental animation effect found here is consistent with Zhang's studies. However, the detrimental effect only manifested on search speed; search accuracy was not affected by animation. This is consistent with our prediction and with previous findings that attention can change the time course of a visual search task without affecting its asymptotic accuracy level [3].

Although color animations made the visual search performance worse when compared to control condition, the extent of the detrimental effect did not depend on animation location or timing. This finding is inconsistent with Zhang's finding. This inconsistency could be explained by the difference in the search screen layout and animation timing between the two studies. There was more space between lines in Zhang's studies, which made the search task easier than the current study. Zhang's studies showed that the detrimental effect of the animation varied with task difficulty. In addition, Zhang's studies were conducted in a web browser while the current study was conducted on a desktop application. Once started, animations in Zhang's studies lasted until the end of the search and different animations had different durations. In the current study, all animations lasted 10 seconds to control the effect of exposure duration on the memory about the animated strings.

The Effects of Animation Location and Timing on the Memory about Animated Strings. Animation location had no significant effect on participants' ability to recognize the animated strings. This is inconsistent with our prediction but consistent with finding that the detrimental effect of an animation did not depend on its location, which means that the amount of attention an animation received did not vary with its location. Animation timing had significant effect on participants' ability to recognize the animated strings. When examined closely, only the difference between the strings animated close to the end of the search and the strings presented in on-off-on animation was significant and the strings animated close to the end of the search were recognized significantly better. One possible explanation for the superior sensitivity in recognizing the strings animated close to the end of search is the relatively shorter interval between the animation and the recognition test. The strings in on-off-on animation were only presented for half of the animation duration every time they appeared. Although the total amount of the time presented was equal to the other conditions, the short presentation interval at appearance might not give the

participants enough time to encode the strings, which could lead to the poor performance in recognition test.

Animations increased the appeal of the animated strings. This is consistent with previous findings on exposure frequency effect [9].

4 General Discussion and Conclusion

The prediction that animations will slow down the search process but will not affect the search accuracy and the prediction that animations will damage visual search performance was supported when the animations were in color. The effects of an animation on search performance did not vary with animation location and timing, which is inconsistent with the prediction from previous research. The prediction that animation location and timing will have no effects on the preference judgment of the animated strings was supported.

As predicted, animations in color were more distracting and damaged primary visual search task more than animations in black-and-white and animations in color were remembered better than animations in black-and-white in recognition test. The current study found that color animations damaged visual search task while black-and-white animations had no significant effects on visual search performance. This is consistent with previous study where brightly colored animations damaged visual search tasks more than dull animations [11]. One reason that color animation is more damaging to visual search than black-and-white animation is that object in color is more salient than object in black-and-white in a black-and-white search screen. Therefore, color animations tend to attract more attention from the participants and distract them more in the visual search than animations in black-and-white.

Although Zhang demonstrated in her studies that animations damaged visual search performance, she did not pinpoint which aspect of the visual search task was affected. The current study showed that animations slowed down the visual search but had no effects on search accuracy.

The most important finding in the current study is that animations increased the likeability of the animated strings even when the primary search task was not affected by the animations and the explicit memory about the animated strings was very poor as in Experiment 1. This is consistent with our prediction and suggests that advertisers can reap the benefits of the advertisement without the user having to sacrifice information seeking performance. This is the ideal win-win situation for the user, the content provider, and the advertiser.

In summary, the current study investigate the effects of animations on visual search tasks on a finer scale by examining the search speed and accuracy separately and by calculating the attentional effects of animation on visual search using signal detection model. We found that animation in color did slow down the visual search but had no effect on the accuracy and animation in black-and-white had no effect on visual search task. Mere exposure to nonsense strings in animation increased the likeability of the strings significantly even when the strings were not recognized.

References

1. Bayles, M.E., Chaparo, B.: Recall and recognition of static vs. animated banner advertisements. In: Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting, USA (2001)
2. Cameron, E.L., Tai, J.C., Eckstein, M.P.: Signal detection theory applied to three visual search tasks – identification, yes/no detection and localization. Spatial Vision 17(4-5), 295–325 (2004)
3. Dosher, B.A., Han, S., Lu, Z.-L.: Parallel processing in visual search asymmetry. Journal of Experimental Psychology: Human Perception and Performance 30(1), 3–27 (2004)
4. Finlay, K., Marmurek, H.H.C., Morton, R.: Priming effects in explicit and implicit memory for textual advertisements. Applied Psychology: An International Review 54(4), 442–455 (2005)
5. Interactive Advertising Bureau. Internet advertising revenue report 2005 (April, 2006) Retrieved 2006, 8, 24, at http://www.iab.net/resources/ad_revenue.asp
6. Interactive Advertising Bureau IAB/PwC release first half 2006 Internet ad revenue figures (September, 2006) Retrieved 2007, 2, 19, at http://www.iab.net/news/pr_2006_09_25.asp
7. Hein, K.: Improve Your Online Marketing, Incentive (November 1997)
8. McGalliard, K.: Animate your web site gif graphics, Editor and Publisher (May 1998)
9. Zajonc, R.B.: Feeling and thinking: preferences need no inferences. American Psychologists 35, 151–175 (1980)
10. Zhang, P.: Will You Use Animation on Your Web Pages? Experiments of Animation's Effects and Implications for Web User Interface Design and Online Advertising. In: Romm, C., Sudweeks, F. (eds.) Doing Business on the Internet: Opportunities and Pitfalls pp. 35–51 Springer-Verlag in conjunction with the British Computer Society (1999)
11. Zhang, P.: The effect of animation on information seeking performance on the World Wide Web: securing attention or interfering with primary tasks. Journal of Association for Information Systems (2000)
12. Zhang, P.: The impact of animation timing and location on visual search task performance in the web environment. In: Proceedings of the Americas Conference on Information Systems (AMCIS'2001), Boston, MA, pp. 1361–1366 (2001)

Using Root Cause Data Analysis for Requirements and Knowledge Elicitation

Zhao Xia Jin[1], John Hajdukiewicz[2], Geoffrey Ho[2], Donny Chan[1], and Yong-Ming Kow[3]

[1] Honeywell Technology Solutions Lab, China
Zhao.Xia.Jin@Honeywell.com
[2] Honeywell Automation and Control Solutions Laboratory, USA
[3] In-Situ Research Pte Ltd, Singapore

Abstract. The purpose of this paper is to present a technique, called Knowledge FMEA, for distilling textual raw data which is useful for requirements collection and knowledge elicitation. The authors first give some insights into the diverse characteristics of textual raw data which can lead to higher complexity in analysis and may result in some gaps in interpreting the interviewees' world view. We then outline a Knowledge FMEA procedure as it applies to qualitative data and its key benefits. Examples from a case study are presented to illustrate how to use the technique. Proposed Knowledge FMEA brings many advantages such as forcing the analysts to become deeply immersed in the raw data, identifying how the information is connected in causation, classifying the data according to why, what, how formulations and quantifying the findings for further quantitative analysis.

Keywords: Root Cause Analysis, Failure Modes and Effects Analysis (FMEA), Thematic Analysis, Qualitative Research.

1 Introduction

Today, user centered design research emphasizes the importance of viewing the world from the stakeholder's point to better understand user needs and to drive design. Mature qualitative techniques, such as interview and focus groups, have been widely used for knowledge and requirements elicitation. Qualitative techniques generate large quantities of raw data from which researchers distill end users' perspectives, knowledge and world view. The large quantities of raw data, which is usually in the format of field notes, transcripts of interviews or focus groups, can be very complicated to analyze due to the nature of the technique utilized and the way stakeholders view the problem or domain. The raw data of qualitative research can be a relatively unorganized knowledge base. In order to uncover the mystery of this unorganized knowledge base, we identify four dimensions of textual raw data that can simplify analysis and aid in interpreting the interviewees' world view: (a) semantic and thematic classification, (b) explicitness, (c) causation and problem solving path and (d) significance and salience.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 79–88, 2007.

Semantic and thematic classification. Qualitative studies usually rely on inductive reasoning processes to interpret and structure the meanings that can be derived from data [1]. Most analysis methods focus only on classifying the meaning of data according to semantic and thematic frameworks. Typical examples of these methods include thematic analysis and content analysis. Semantical content analysis [2] deals with semantic units (e.g., keywords) of the communication content. Thematic analysis deals with thematic categories (called themes) which are usually semantic entities that have been selected as having a special meaning. Though it is often not explicitly claimed as the method of analysis, Braun and Clarke [3] argued that a great deal of analyses are essentially thematic but is either claimed as something else or not identified as any particular method at all. These methods of analysis are indeed useful but may be limited in extracting more details in tacit knowledge such as interconnections of why and how events occur.

Explicitness. Knowledge is different in the degree of explicitness. Explicit knowledge is transmittable in formal, systematic language. However, tacit knowledge is personal, deeply rooted in a specific context, and hard to formalize and communicate [4, 5, 6, 7]. Commonly claimed notions say, "we know more than we can tell" [4] and "users do not know how to articulate their real needs and requirements". For example, the interviewee may be able to express the information in knowing-that but not in the form of knowing-how. Therefore, the large amounts of raw data consist of unstructured explicit knowledge, tacit knowledge and have complex relationship between data. The difficulty arises because we are trying to translate the tacit knowledge into explicit or codified knowledge. The difficulty is magnified because we attempt to externalize the complex relationship between data.

Causation and problem solving path. Many researchers point out that data analysis should take into account the interrelationship between the themes as they occur within the data [8, 9]. Causation, as an important data relationship, is often overlooked in discussions of qualitative research [10]. Causation is the relationship between cause and effect that holds association between events, properties, causes, influence, determination, contribution, effects, and so on. Cause-effect relationship is the primary aspect that ties into how to prevent the problems from failure again. Baptiste [10] suggested that the analyst should ask questions to reveal the causation but he did not provide an effective method to conduct the analysis. Another exception is pragmatically content analysis, which includes procedures for classifying signs according to their probable causes and effects and emphasizes said the purpose of discourse [2]. The causation and its importance in problem solving drive us to analyze data according to cause-effect relationships and problem solving paths, which involves identifying problems, evidence, root causes, consequences, and providing solutions as corrective and preventive actions.

Significance and salience. Though the usual focus of data analysis of qualitative research is on the meaning of the information collected, quantitative analysis can be followed to uncover the significance and salience of the findings. For example, the quantitative dimension of the content analysis, centers on determining the frequency of keywords or phrases in context [11, 12]. Another example is to simply count the

frequency of occurrences of each of the themes determined [13]. Such analyses are useful for comparing the relative frequency, and arguably the salience, of items of analysis but do not tell us much about the patterns among the items [14]. Also, such analyses involve only simple forms of quantification mostly in the format of frequency but do not consider the severity of the problems which are rooted in the meaning of the textual raw data. Ultimately, "all qualitative data can be coded quantitatively since anything that is qualitative can be assigned meaningful numerical values" as pointed out by Trochim [15].

As summarized by Guest and McLellan [8], during the past few decades, researchers have developed a host of methods for structuring and analyzing textual data (for a review of these, see Dey 1993 [16]; Bernard and Ryan 1998 [17]; LeCompte and Schensul 1999 [18]). Most of the methods only focus on analyzing certain dimensions of knowledge and seldom detail the analysis procedure on how to address the explicitness and root cause paths of the information. This paper applies Root Cause Analysis as a method to analyze the interview data and generate a Knowledge Failure Modes Effects Analysis (FMEA).

2 What Is Knowledge FMEA?

Knowledge FMEA is a FMEA-like data analysis technique designed to decompose and organize the textual data according to cause-effect relationships, quantify the findings by Severity, Occurrence, Detection and Risk Priority Number (RPN) ratings, inform features, functionality and possible solutions ideas. Knowledge FMEA addresses the four dimensions of textual raw data mentioned earlier. Here, it forces the analysts to be very clear about how the information is connected (why, what, and how formulations) and separate the data into root cause and problem solving path.

3 Procedure of Doing Analysis with Knowledge FMEA

The six stages of doing analysis with Knowledge FMEA are summarized in Table 1.

3.1 Get Familiar with Data

Immerse oneself in the data and get a sense of the data as a whole before breaking it into parts. Data collected from an interview might be written information (interview notes, field observational notes) and audio/visual information (audio recordings, video recordings, pictures) that might need transcribing. The analysis begins with getting familiar with the data by repetitious reading to search for the underlining meanings and patterns. Perusing the data allows the analyst to make use of prior knowledge about the data collected and instill initial analytic thought as a whole before breaking it into parts. During this familiarization process, major themes begin to emerge. Making additional notes, color codes and marking initial categorization will be useful for latter stages of coding and will provide a reference point in the later stages. Subsequently, the data can be coded into short statements which will be grouped into the different themes that house them latter.

Table 1. Procedures of Knowledge FMEA Analysis

Analysis Stage	Procedure
State 1: Get familiar with data	(1) Transcribe data (if necessary). (2) Repetitious reading to get familiar with the data and get a whole picture. (3) Code data where the themes start to emerge. (4) Make notes and mark of the initial idea.
State 2: Fundamental Thematic analysis (optional)	(1) Complete coding process with an iterative process. (2) Structure themes and concept map of themes. (3) Classify and group the statements. (4) House the statement according to standardized themes.
State 3: Root cause Analysis	(1) Tailor fields of root-cause path. (2) Conduct root cause analysis iteratively. (3) Standardize the analysis results if necessary.
State 4: Quantitative rating	Rate Severity, Occurrence, Detection and calculate RPN (Risk Priority Number).
State 5: Review the analysis	(1) Review the analysis by self revision or peer review. (2) Follow up or verify with participants if necessary.
State 6 : Quantitative analysis & interpretation	Perform statistical analysis using quantitative RPN data and interpret the analysis.

3.2 Fundamental Thematic Analysis

Optional step to classify the data for further analysis. Thematic analysis begins when you start coding the raw data, breaking them into parts and grouping them by proximity. A quick way to describe a phenomenon, identify and verify the patterns from unorganized data is to classify and group the data. A good method to do this is to use thematic analysis as a fundamental step to classify and organize data into themes.

How to use thematic analysis is widely reported in literature [3, 15]. Thematic analysis can be done in a deductive or inductive manner. In the deductive (top-down) method, data analysis is determined by the research objectives and the themes pre-exist in theory or literature. While in the inductive (bottom-up) method, the themes are developed from the empirical data or emerging from reading and re-reading the raw data. Themes should be refined, decomposed and structured so that a concept map (e.g., hierarchy chart) of themes can be created.

In general, thematic analysis helps to select, focus, abstract and transform the data collected into manageable information segments to show patterns. This step lets analysts to familiarize the data further, organize the statement and better prepare analysts to perform further root cause analysis. House the statement according to standardized themes.

3.3 Root Cause Analysis

Next, root cause analysis is conducted to further analyze and understand cause-effect relationship with the statements categorized earlier. A convenient way is adopting existing mature tools like FMEA. While FMEA can be very complex, not all the fields of root cause path (e.g., failure modes, causes, consequences and corrective actions) need to be adopted. These fields shall be tailored to fit the research interest and the level of detail a researcher trying to identify.

Generally, in order to elicit more knowledge, researchers actively ask open questions (e.g. What, When, Where, Why, How) and use probing technique on participant statements. As a result, researcher notes are a collection of answers that

are scattered and do not have coherent relationship. To piece this part and parcel together to tell a story based on cause-effect relationships, researcher continue with further iteration of active reading and asking what, why and how on the collected statement. This process is consistent with widely reported root cause analysis method. By doing so, assumptions might be made on the statements to uncover the relationship. When statements collected are ambiguous, they can be left unclassified with additional notes or markings. Both assumed and uncategorized results need further clarification and verification from participant or peer review to solve the problematic analysis and at the same time eliminate discrepancies.

3.4 Quantitative Rating

This step involves rating all statements after they are categorized and cause-effect relationships are identified. In this step, the analyst needs to rate three items adopted from FMEA analysis tool including 1) Severity (S) – how serious the problem is, 2) Occurrence (O) – how frequent does it happen and 3) Detection (D) – how easily it can be detected. By multiplying these three ratings, a Risk Priority Number (RPN) will be generated that reflects the impact of the happening. These FMEA tool's rating provide a better representation of the statement in a quantitative manner from different angles. For example, RPN (9) = S (9) ×O (1) ×D (1), explains a finding that is serious but might seldom occur or might be easily detected. Another example is, RPN (9) = S (5) × O (9) ×D (5), explains a moderately serious problem that always reoccurs and is oftentimes detectable. The combination of these three ratings creates a high RPN that means it brings high risk and big impact that immediate attention is needed.

Effective and well defined standard rating scales should be used; commonly we use rating scales of 0, 1, 3, 5, 7, 9. Different rating scales might be used separately for Severity, Occurrence and Detection. However, these usually have been standardized in a mature FMEA tool. These S, O, D and RPN provide additional insight that helps to interpret data. The rating process can be performed by a few researchers and can take place at a different time. Results can later be easily shared and combined.

3.5 Review the Analysis

Self review or peer review of the thematic analysis, root cause analysis, and rating help to identify and eliminate discrepancies that arise. Researcher might want to reconsider some categorization that seems to be ambiguous, especially on those results that had been marked with notes or labels. Peer review share more insight of the respondents' world view according to the experience during the study. Any discrepancies found during reviewing process, need be resolve and achieve common agreement among team member. Researcher might need additional follow up with participants and verify the problematic results. Having a reviewing process helps to generate a more reliable (e.g. check and balance by peers) and a more valid data (e.g. standard rating and eliminate discrepancies).

3.6 Quantitative Analysis and Interpretation

With the S, O, D and RPN rating score, further quantitative analysis can be conducted such as descriptive statistic analysis. For example, a matrix with two theme categories

can be created to identify the areas with higher/lower scores. Also, a matrix can be created to analyze the correlation between two themes. Researcher can also compare topics/areas within a sub-theme separately. Finally, these analyses can be interpreted together with qualitative profile and patterns. Refer to examples of Case Study for the more details.

4 Advantage of Using Knowledge FMEA Analysis

i) Researchers are forced to be explicit about the relations in the data. The root-cause analysis forces the analysts to deeply immerse themselves in the raw data and to be very clear about how the information is connected in causation. The method identifies the information that is explicitly or implicitly expressed by the interviewees and thus reflects the interviewees' world view more accurately and with greater detail.

ii) Facilitates problem solving. The analysts are forced to classify the data according to why, what, how formulations and problem solving paths, and thus facilitates problem solving and informs features, functionality, and solutions.

iii) Enables zooming results to areas of interest. The researchers will be aware of the assumptions that they made during analysis and what needs to be followed-up or verified.

iv) S, O, D and RPN ratings quantify the findings for further analysis. These numbers can be further analyzed to achieve greater insight into the meaning of the data; for instance, ratings can provide further prioritization of information or help to examine specific hypotheses.

v) Knowledge FMEA analysis is flexible. The fields that need to be analyzed are customizable and tailor-able depending on the level of detail that the research needs. Also, knowledge FMEA can work in conjunction with other qualitative analytic methods, such as thematic analysis. Finally, even if you just analyze part of the statements stemming from the raw data, the analysis method still provides a deep level of insight.

5 Examples of Case Study

Below is a case study where we used the Knowledge FMEA process.

5.1 Project Background

This field study was a multiple stakeholder study (15 participants from five different stakeholder categories were interviewed) and involved multiple sites (3 different sites). The purpose of the project was to identify the key factors that affect the productivity of work of a complex collaborative project environment in various areas, such as activities, environment, tools, communication and coordination process. Semi-structured contextual inquiry and observation were used to elicit knowledge from participants. For each interview, two researchers conducted the interview as a team.

5.2 Organizing Raw Data and Fundamental Thematic Analysis

As all site visits happened in different stages, a preliminary deductive thematic analysis took place immediately after the interview and produced a report for each interview. The process decomposition of the life cycle of the work was used as a theme category to classify the statements. Also, signs were used to mark the comments (positive, negative, neutral comment) in the reports. This initial familiarization and categorization helped to further improve the site visit design and provided preliminary organized data.

Subsequently, statements categorized in the preliminary report were combined and transferred to a spreadsheet by inserting individual statements into rows. Another new theme category (in here is 'Productivity Area' of from theory [19] for the case study) was added to extend the level of granularity. For the root cause analysis, interested fields from an FMEA (e.g. failure mode, root causes, correcting actions) were tailored into the analysis template. We customized the field names according to the research interest. For example, the term X-Factors replaced the original term, Failure Mode; X-Factors are the known and potential factors (include failures) that could increase, prevent or degrade the productivity, quality and efficiency. Causes of Failures were renamed to Diagnostic Causes referring to any known and potential causes. Correcting Actions was changed to Possible Solutions as we more concerned with the design changes which can correct the failure modes and improve the efficiency of the tools. Also, rating areas such as S, O, D and RPN, were selected. All these fields were combined and form an analysis template as in Table 2. All the statements were coded and analyzed based on this template.

Table 2. Analysis template with two deductive theme criteria and root cause path

Thematic categorization		**Root cause analysis**							**Original Statement**
Process/Sub -function	Productivity Area	X-Factors	Severity	Diagnostic Causes	Possible Solutions	Probability of Occurrence	Detection	RPN	Original statements

5.3 Root Cause Analysis

As mentioned, root cause analysis was conducted to identify or elicit certain root causes behind a problem and ideas to prevent the problems. In Table 3, an inductive example (example 1) and a deductive example (example 2) of conducting root cause analysis are shown by applying the process of asking what, why and how

Statement 1: "There should be no XXX step anymore in HW engineering...This should be automatic...The biggest challenge is if you make a mistake in XXX step, it's about 3 to 5 times the cost to change the mistake..."

As shown in Row 1, the participant's statement explicitly states that he was referring to a step in the 'HW Engineering' process and that the step was 'poorly supported by Tool Features'. This problem resulted in 'about 3 to 5 times the cost' due

to 'no automation support'. Apparently, this problem can be improved by 'automating X steps in the tool'.

Statement 2: "[We] often go to training and still don't know enough..."

As shown in Row 2, a participant provided feedback of the training they had attended. By just looking at the transcript, various interpretations and assumptions can be summarized about the training. Assumptions were made and translated as 1) product is too complex and difficult to understand, 2) training hours are not sufficient therefore need longer hours, or 3) poor training quality. Based on the interview context, assumptions were made that the participants felt that the training was 'poorly organized', was 'without sufficient information about the product' and had 'no clear training objectives'. Through brainstorming, two solutions, 'setup well organized training' and 'provide sufficient information supported by tool' were provided. Researchers then verified their assumed statements from the participant and confirmed that these were the actual causes of their discontent.

Table 3. Examples of analyzing path of root cause analysis

Process/ Sub-functions	Productivity Area	X-Factors	Root cause	Potential Solutions
HW Engineering	Tools	Poor support of XXX step	[1] No automation support [2] Cost about 3 to 5 times the cost to change the XXX mistake.	Automate XXX
General	Knowledge & skills	Poor product training	[1] Insufficient information about the product. [2] Training poorly set up [3] No clear training objective	Setup up well organized trainings Provide sufficient information supported by tool

Finally, an affinity tree was created by simply sorting the spreadsheet file according to different themes. This helped to bring out the coherent relationship of the results and tell a story by linking the what, why, how formulation. As well, it demonstrated the consequences. We also standardized the X-factors to see patterns created by X-factors.

5.4 Qualitative Analysis and Data Interpretation

After all the statements were categorized according to themes and the root cause analysis was completed, two researchers rated Severity, Priority, and Detection according to a standardized rating scale to generate RPN ratings. Accumulated experience in the field and further understanding during the process of familiarization and categorization helped the researchers rate these statements more accurately. In peer review session, researchers shared their insight and opinions about the statements. At times there were debates to resolve discrepancies in their opinions, particularly when the RPN scores had large inter-rated deviations.

Table 4. Example of matrix table with two theme categories

Process/Sub-Functions Productivity Areas	Productivity Area 1	Productivity Area 2	Productivity Area 3	Productivity Area 4	Productivity Area 5	Productivity Area 6	Productivity Area 7	Productivity Area 8	Productivity Area 9
Process/Sub-Functions 1	34	20	65	6	31	33	31	101	28
Process/Sub-Functions 2	13	26	1	/	/	/	/	23	9
…	…	…	…	…	…	…	…	…	…

Table 4 illustrates a cross-combination of RPN scores to compare main two dimensions. One dimension reflected the work process and their sub-functions and the second dimension reflected key productivity themes or areas). Certain cells were empty because no data was collected for that theme. These empty fields also created a check and balance process where the data collected has to correspond to the role and job responsibilities of the participants.

The S, O, D, and RPN could be easily sorted in the spreadsheet to show the most serious problems, the problems that most frequently, the problems that were not easily detected, and the impact of the overall problem.

6 Conclusion and Discussion

The present paper identified four dimensions of textual raw data that contributes to the complexity of qualitative analysis. These dimensions oftentimes will result in gaps in interpreting the interviewees' world view. The authors focused on the cause-effect relationship and introduced a root cause analysis technique that can be applied to qualitative data analysis. As we have shown throughout this paper, Knowledge FMEA analysis is not a complex method and has many advantages. Used properly, Knowledge FMEA can be both a useful and a powerful technique, but its use must be associated with in-depth knowledge of the raw data. A significant drawback is that it is time-consuming if you try to analyze large amounts of data with this technique. However, you'll get results are easily interpretable and still rich with details. Ultimately, one's approach will depend on the type of data, the granularity of coding, and the research objectives, as Greg Guest [8] pointed out.

References

1. Thorne, S.: Data Analysis in Qualitative Research. In: Evidence Based Nursing, vol. 3, pp. 68–70. University of British Columbia Vancouver, Canada (2000)
2. Janis, I.L.: The Problem of Validating Content Analysis. In: Lasswell, H.D., Leites, N.C. (eds.) Language of Politics; Studies in Quantitative Semantics, pp. 55–82. MIT Press, Cambridge, MA (1965)
3. Braun, V., Clarke, V.: Using Thematic Analysis in Psychology. In: Qualitative Research in Psychology, vol. 3, pp. 77–101. Edward Arnold, United Kingdom (2006)
4. Polonyi, M.: The Tacit Dimension. Doubleday & Company, Garden City (1996)
5. Crowley, B.: Tacit Knowledge, Tacit Ignorance, and the Future of Academic Librarianship. College and Research Libraries 62 6, 565–584 (2001)

6. Gourlay, S.: Tacit Knowledge, Tacit Knowing or Behaving? In: Third European Conference on Organizational Knowledge, Learning, and Capabilities, Greece (2002) at http://www.alba.edu.gr/OKLC2002/Proceedings/pdf_files/ID269.pdf
7. McInerney, C.: Knowledge Management and the Dynamic Nature of Knowledge. Journal of the American Society for Information Science and Technology 53(12), 1009–1018 (2002)
8. Guest, G., McLellan, E.: Distinguishing the Trees from the Forest: Applying Cluster Analysis to Thematic Qualitative Data. In: Field Methods, vol. 15, pp. 186–201. SAGE, Thousand Oaks, CA (2003)
9. Henderson, C., Yerushalmi, E., Heller, K., Heller, P., Kuo, V.H.: Multi-Layered Concept Maps for the Analysis of Complex Interview Data. In: Roundtable Discussion presented at the Physics Education Research Conference, Wisconsin (2003)
10. Baptiste, I.: Qualitative Data Analysis: Common Phases, Strategic Differences, Forum: Qualitative Social Research, 2(3) (September 2001), at http://www.qualitative-research.net/fqs-texte/3-01/3-01baptiste-e.pdf
11. Berelson, B.: Content Analysis in Communication Research. In: Glencoe, I.L.: The Free Press Publishers (1952)
12. Kaplan, A.: Content Analysis and the Theory of Signs. In. Philosophy of Science 10, 230–247 (1943)
13. Leites, N.C., de S. Pool, I.: On Content Analysis. Library of Congress, Experimental Division for Study War-time Communications. Documents 26, 1–2 (1942)
14. Richards, T., Richards, L.: Using Hierarchical Categories in Qualitative Data Analysis. In: Kelle: Computer-Aided Qualitative Data Analysis: Theory, Methods, and Practice, pp. 80–95. Sage, Thousand Oaks, CA (1995)
15. Trochim, W.M.: The Research Methods Knowledge Base, 2nd edn. (2006), at http://trochim.human.cornell.edu/kb/index.htm
16. Dey, I.: Qualitative data analysis: A user-friendly guide for social scientists. In: Bernard, H.R., Ryan, G. (eds.) 1998 Routledge, New York (1993)
17. Bernard, H.R., Ryan, G.: Text analysis: Qualitative and quantitative methods. In: Bernard, H.R. (ed.) Handbook of methods in cultural anthropology, pp. 595–645. Altamira, Walnut Creek, CA (1998)
18. LeCompte, M.D., Schensul, J.J.: Analyzing and interpreting ethnographic data. Altamira, Walnut Creek, CA (1999)
19. Accel-Team: Productivity Improvement (2006), at http://www.accel-team.com/productivity/productivity_01_what.html

What Stories Inform Us About the Users?

Yong Ming Kow[1], Angela Tan[2], and Martin Helander[2]

[1] In-Situ Research Pte Ltd, 32 Maxwell Rd, #03-07, S(069115), Singapore
yongming.kow@insituresearch.com
[2] Nanyang Technological University, School of Mechanical and Production Engineering. Nanyang Technological University. Singapore 639798
bubblegenie@pmail.ntu.edu.sg,martin@ntu.edu.sg

Abstract. Storytelling has been used to elicit subconscious schemas that were formed from user experiences. Subjects were required to construct a working model using external and internal sources of information. Known methods of performing narrative analysis included Projective Tests, Narrative Analysis, and Cognitive Tasks Analysis. Three studies using storytelling methods were done with these methods. One with older adults, two with kins of older adults, and three with refinery operators. In the case of older adults, users were asked to make up stories for a fictitious person to extract cultural norms and knowledge. For the refinery study, we can look into real stories for more reliable data. Many types of design information were extracted: (1) emotional and functional needs, (2) functional dependencies, and (3) expertise. Arguably, these information will be hard to come by if a direct interview technique is conducted. Reasons due to the richness of information embedded in stories.

Keywords: Storytelling, Projective Test, Cognitive Task Analysis, Interview Method.

1 Background of Story Telling for Requirements Elicitation

Narration has been used since ancient times [1]. Despite the amount of work put into ordering the world into distinct variables and behavior, interestingly, narratives remains indispensable. An example is the paper you are reading! Some authors suggest that human thinks in a narrative fashion, in a chain of events in a chronological order [1, 2]. Arguably, narratives contain richer informational contents which variables cannot portray. This paper attempted to review lessons learnt from three cases of using storytelling method to identify design requirements, and what we can learn from them.

1.1 Extracting User Needs from Stories

Storytelling have been used to extract user needs. User needs may be in a form of basic needs, motivation, goals, and their relationship with people and artifacts found within the community.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 89–97, 2007.

A form of studying true stories within a communal settings, known as Narrative Analysis, has been used in social ethnographic studies of culture and social issues. These were frequently done in situation where sample size was too small for quantitative analyses [2].

Another form of storytelling, of fictitious but socially valid stories, is the Thematic Apperception Tests from the family of Projective Techniques. In marketing, it was used to understand users' needs [3]. In one study, pictures of cheap versus expensive cosmetic were used. Users subconsciously see the cheap cosmetics as "greasy and oily" and the expensive cosmetic as leaving the skin "clear, refreshed and young-looking." In another study in 1950s, pictures of "housewife who used instant coffee" versus "housewife who used ground coffee" were compared. The former was viewed as "lazy" while the latter as "thrifty." This impression disappeared in a replicated study done in 1970s.

Using fictitious stories may imply loosing external validity of the findings. However, if a product is truly innovative, it is difficult to find any users with prior use experience. What people claim to be needs may merely be false assumption based on social desirability biases. In addition, some potential users may not be aware of that they have certain needs.

1.2 Extracting Tacit Knowledge from Stories

Stories may also be used to elicit tacit knowledge through revisiting certain parts of a story a multiple levels of details. This is known as "Sweeps" in Critical Decision Method (CDM) [4]. This was most effectively used to probe expensive or life threatening incidents that required good judgment from expert users (known as Subject Matter Experts in CDM). Since incidents cannot be simulated, interviews were conducted to investigate recent cases of interest. Users were asked to narrate what happen and important chain of events were revisited in several level of details.

CDM utilizes four "sweeps" [5]. In Sweep 1, the Subject Matter Expert (SME) selects a challenging incident that recently occurred. He describes the chronological sequence of the incident in Sweep 2 and progressively deepens the story in Sweep 3. Finally, the SME discusses the expert-novice difference using what-if scenarios in Sweep 4.

2 Theories Substantiating User Requirements Found Beneath Stories

Stories contain rich information about users and their environment. The projective hypotheses state that stimuli from the environment are perceived and organized by the individual's specific needs, motives, feelings, perceptual sets, and cognitive structures, and that in large part this process occurs automatically and outside of awareness [6]. Users were thus guided unknowingly by a set of rules known as Schemas. In storytelling, users and their environment were reverse engineered out of such schemas by consulting external and internal sources of information [7]. In constructing stories, many subconscious processes were utilized, explaining why direct interviews cannot be used to elicit the same types of information.

2.1 Schema Theory

Schema theory predicts that people are aware of what to do next in familiar circumstances as if these are autonomic reactions. Schema theory has similarities with RPD (recognition primed decision making) in human factors, decisions for which action alternatives are derived from recognition of critical information and prior knowledge. Human uses intuition in what they are doing well [8]. Klein [1] further explained that intuition allowed experienced firefighters to see a new situation as typical, thus automatically generating a course of action. This is unlike inexperienced firefighters, who may be lost when facing new situations. Schemas may be applied to stereotypical identities associated with groups, such as information about social positions and stratification statuses, such as gender, race, age, or class [9].

2.2 Dual Theory of Information Processing

Dual theory of information processing claims that human process information using both rational and experiential systems of thought [10]. The rational system is deliberate, effortful, and analytical, while the experiential system is emotional, autonomous, and subconscious [7]. Dissociation between rational and experiential thinking is evident in hedonic psychology. Berridge [12] noted that a person find an object attractive (like) yet denying the desire to own it (want). For example, recovering drug addicts could, through the experimental system, actively seek drugs without consciously being aware of such tendency. In TAT, it is believed that narratives can be used to assess implicit needs or “likes” though indirect reporting [7].

2.3 Information Contents Within Stories

According to Klein [1], stories contain information such as agents, predicaments, intentions, actions, objects (tools), causality, context, and surprises. In analyzing user activities, Carroll and Rosson [13] observed that stories contain the following elements: setting, actors, task goals, plans, evaluation, actions, and events. Similarly, in Thematic Apperception Test, stories are believed to contain: dilemma, intention, complications, means, and outcome [7]. Besides information content, mediation between elements within stories reveal: causal relationships, temporal dimension, and themes or lessons behind each story, states of the objects, chronological order of events and actions, actors, causal relationships between pairs of states where each pair can be consider as events, and actions transforming states [1], [2], [11].

2.3.1 Knowledge Within Stories

One of the key strength of studying narrations is in identifying user knowledge. There are two types of knowledge: cultured laws and learnt expertise [2].

Cultured Laws. At noon time, most Chinese would inevitably think about lunch. But a question posed by the author to some Americans asking, “When is lunch time?” was returned with the question, “Are you hungry? We can get a sandwich!” Sandwich is not the author’s concept of lunch. In any socializing situation, dining is the standard Chinese activity that binds people in a conversation. These is a case of cultured laws

governing our lives. Narrations can act as a valuable window to understand social laws and motivations of actors [2].

Learnt Expertise. Stories can uncover learnt expertise but some of which are deeply hidden away from user consciousness. Such expertise, when uncovered, can be valuable to training or aiding of newer users, or automation of knowledge. Rasmussen uncovered three levels of learnt expertise: Rule-based expertise (If/Then), knowledge –based expertise (analytical), and tacit expertise [14]. Tacit expertise are resistant to being articulated, and is often used subconsciously due to pattern recognition or perceptual motor feel [14]. Tacit expertise are the prime candidate of storytelling technique which differentiate itself from direct interviewing methods.

3 Case Studies

Three separate studies were conducted using storytelling.

3.1 Study 1: Identifying Needs of Older Adults Using Innovations

The objective of this study is to identify the needs driving usage of technologies by older adults. Identified technologies were either very new or unavailable. Therefore, the technique has to cater for future scenarios based on socially valid scenarios. The researchers were interested in what cultured laws and personal interests will be involved in technology usage.

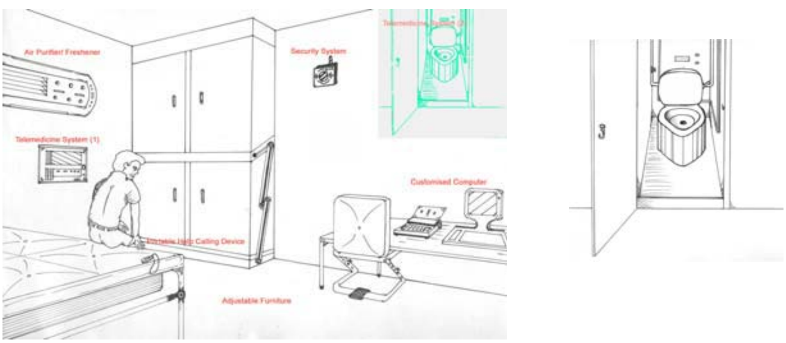


Fig. 1. Illustrations of stimuli used in the projective tests

Sixteen older adults were interviewed at their home. Their characteristics varied widely. Pictures of a person using several home technologies were shown to each older adult, see figure 1. Care was taken that the picture was sufficiently neutral and do not lead the users into predetermined conclusion. This was followed by a question of “how is the man using the device?” After the users had provided a few examples of using a tool, it was sometimes possible to probe the circumstances surrounding

examples. Questions asked included: “Who is he,” “what kind of person is he,” “what sickness does he have,” “where is he going,” “how old is he,” and “why does he use it in this way?”

Altogether sixteen residents were interviewed. Some were staying with their spouse or children under fifty-five years old. They were mainly middle class. Many were in touch with their children who might give them allowances periodically. Their characteristics varied widely.

Two categories of these data were observed: (1) usability concerns and (2) projected needs. Usability concerns were elements that would enable the products to be more accessible. These were features such as “to be used by children,” “fear of using due to lack of knowledge,” “to use without being literate,” “using conversation to communicate (with sales assistance in teleshopping),” and “to pay (the tools) by installment.” There were also projected objects, persons, or characteristics of which users associated with the products in order to come to their conclusions. For example, “(the teleconferencing system) is costly and not for family,” “person (who uses teleshopping) is old and uses cane.” For the sixteen users, the total number of different needs registered were one hundred and eighty-six.

3.2 Study 2: Identifying Relationships Between Functional Needs of Older Adults

The objective of this study is to identify the causal relationships between smart home functions. Although seven functions were identified out of a previous card-sorting exercise, its unknown how, in the context of activities, these functions would become dependent or overlap with each other in needs satiation. In this case, a third person interviewing technique was used to tell stories of older adults. Since real activities were warranted, stories told were real accounts.

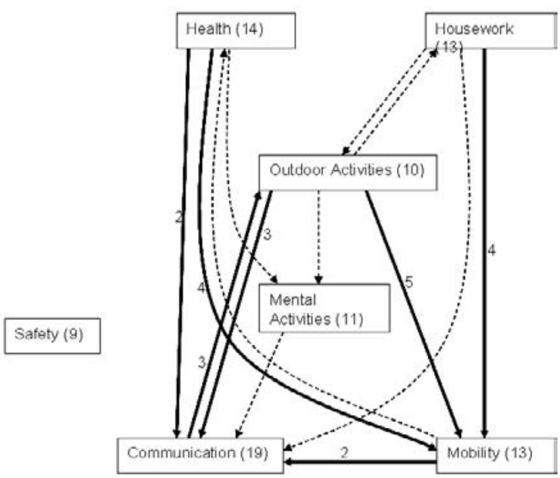


Fig. 2. Causal structure of smart home functions

Nineteen university undergraduate students, eleven males (58%) and eight females (42%), were interviewed. They were in touch with their kins, all above 60 years of age, of which seven were males (37%) and twelve were females (63%). They met their kins at least once every month for the past three years. The subjects were told to write about their kin based on their true experiences [7]. The story should comprise of a series of connected events. The entire story must contain at least one of each seven functions of smart home, namely mobility, communication, safety, housework, mental activity, outdoor activity, and health. But there was no limitation of the number of times each might appear.

A causal structure between functions was identified as shown in figure 2. Each type of relationship, for example communication function causing outdoor activity, outdoor activity causing mobility, or health causing mobility needs to arise, was counted separately. The number beside each arrow shows the number of occurrences in each pair of causal relationships. Many of the relationships were weak with only one case.

3.3 Study 3: Identifying Tacit Expertise Required in Solving a Refinery Incident

The objective of this study is to identify tacit expertise used by a team of refinery operators in solving a particularly refinery incident. Process operations within a refinery is highly team work based. A team of operators worked in differing roles as leader, computer operator, and field operator. A leader is in charge of the entire operation and make major decisions. A computer operator monitors the sensor readings and process conditions. A field operator walks into the plant, monitor the physical instruments and does manual operations. Eight operators in differing roles experience were asked to recall the same incident of how they go about solving the problem.

To explain how things work, the operators tell the story of actions and explained the rationale of these actions [15]. These stories were expressions of the operator's expertise. In order to identify key expertise, the operators were asked to revisit certain parts of the incident. For example, one of the field operator identified a leaked pipe. In an attempt to find out how he did that, he was asked to revisit and deepen that part of the incident. When the operators have difficulties deepening stories, probes were used. Some of these probes were such as: "What were you seeing, hearing, smelling….," "Were you reminded of any previous experience," and "How long did it take actually to make this decision?" [5]. Finally, the operators were asked if they are doing things differently from a novice operator. If so, how is it different. Such as a deepening attempt to elicit critical skills from stories.

4 What Stories Can Tell Us While Direct Questioning Methods Cannot

In some situations, storytelling technique is superior over direct interviewing methods. In direct questioning method, users are asked for their comments on certain issues, such as, "Why do you buy a cellphone?" This question was asked by one of

the master student and he got pretty homogeneous response such as: to make calls, to call someone, to use in emergencies, and so forth. However, there is a good chance these are merely what the users 'believe' and not their real needs. Following are some knowledge which stories are found to elicit but not in direct questioning.

4.1 Personas and Social Undesirable Biases

While most direction questioning techniques were directed at the interviewee, storytelling is especially useful in identifying social norms and realities within the society. In study 1, the response towards a telemedicine device was two widely different personas. In one, the user was "literate" and had "knowledge of computer." In another, he was "old, aching, and walked on cane, and there was no one to help him." Similarly, he could also be sick and "could not walk." Such social realities were inevitably elicited from real experiences surrounding the subject and cannot be dismissed lightly. They can be a start to identify cultural factors that impact on product usage.

Another advantage of third person approach naturally used in storytelling is that users will be forced to use some aspects of their tacit knowledge, which are otherwise difficult to articulate due to reasons such as habituation and social undesirability [1], [7], [8], [16]. For example, people overestimate their willpower, and underestimate their desire to spend, vulnerability to social pressure, and sexual desire [17] Social or personal undesirability means that the respondents reply in a manner that creates positive impression of themselves [16]. Although not all interviews contain social-desirability bias but they were known to be present in socially sensitive issues. These include, but are not limited to: causal factors of charity donation, personal willpower, desire to spend, vulnerability to social pressure, sexual desire, job satisfaction, and sales effort [18], [19], [20]. Nonetheless, since the focus of storytelling is on the story and not on the person, there may be less reporting of personal feelings, especially those related to the character of the person.

4.2 Needs, Goals, and Context

Users inevitably brought up design factors assumed upon the product being discussed. This is because stories forced the storyteller to piece together a coherent and chronological sequence of events which will include the contexts of use. In study 1, usability and projected needs were unintendedly identified. The richness of data that was uncovered by the method provided an opportunity for its complimentary use of SBD (scenario-based design). Storytelling as third person could act as an early endeavor by designers to analyze how products might be used by users, what needs they satisfy, and consider the various design factors to construct usage scenarios for development purposes. From the narration of a culturally ingrained individual, scenarios would be more concrete and realistically bound to real world experiences.

In study 2, one student recalled the scenario where his grandfather had a heart attack at night, and was admitted into the hospital. Due to boredom in hospital, he developed the habit of carrying a cellphone. The cellphone also make it possible for him to assume several of his roles from the hospital bed. From a simple story, we identified the needs, goals, and contexts surrounding a communication technology (cellphone).

4.3 New Themes and Knowledge

Due to the larger degree of freedom for the storytellers, and the challenge of sewing a coherent picture, the storyteller has to inject new sub-themes from cultural personas, actions, artifacts, and cognitive themes to complete the story. Unlike in direct questioning, where the themes are pre-fabricated, interviewees are less likely to bring up unobvious themes. Also the analyst will tend to follow their preconceived thoughts and not venture out to other themes.

In study 3, many operators began with very generic story surrounding the refinery incident. For example, one operator started by saying he heard and saw a leaking pipe in a large multi-level and noisy plant. Only through detailed probing did the interviewer found that he first felt an amount of unnatural vibration in one part of the plant, which prompted him to stop, look and feel around the location. Its only because of his proximity to the leak that he is able to hear the hissing sound and found the leakage. If he had not been an experience operator, he would not be accustomed to the amount of natural vibration within the facility. The location of leakage vibrated more than normal, which first aroused his suspicion.

In order to encourage the insertion of additional themes, interviewers have to encourage the storyteller by suitable probes that trigger the deepening of storylines.

4.4 Activity Scenarios

While products are often designed in accordance to functional needs, these functional needs seldom take into consideration of its relation to the activity it is supporting. Identifying how artifacts mediate within an activity is important to product design. Due to the tight progression of events within stories, there are often enough elements to construct activity scenarios.

In study 2, mobility was found to support most of the current functional needs within the home. It was an end by itself only in one case. But it was the mean to other functions in fifteen cases. Thus enabling the mobility of older adults arguable free them to perform other tasks of daily living. If an older adult has the facility to travel to the nearby clinic, telemedicine devices are being challenge for their usefulness. Similarly, communication function was to be the end in only three cases, but the mean in nine cases. Due to storytelling technique, communication and mobility were found to be the top two enablers of older adults daily living.

As a storyteller develops a story, the interviewer is free to probe other themes relevant to her study. For example, other variables important to construct activity scenarios were personas, artifacts, goals, and cultured laws.

5 Conclusion

Many storytelling methods are still being evolved and we examined only three different methods and their advantages. In these methods, only the principle of having the users constructing story remains, but variations were applied to the technique so that from the stories, the interviewer can achieve her information needs. Some of the variations include, first versus (3) third person perspective (1/2), high level scan (1/2) versus deepening stories (3), and fictitious (1) versus actual events (2/3). Each method has its own strengths and weaknesses and they require further refinement.

References

1. Klein, G.: Sources of Power: How People Make Decisions. The MIT Press, Massachusetts (1999)
2. Abell, P.: Narrative Explanation: An alternative to variable-centered explanation? Annu. Rev. Sociol. 30, 287–310 (2004)
3. Schiffman, L., Bednall, D., Cowley, E., Cass, A., Watson, J., Kanuk, L.: Consumer Behaviour, 2nd edn. Prentice Hall, Australia (2001)
4. Klein, G.A., Calderwood, R., MacGregor, D.: Critical decision method for eliciting knowledge. IEEE Transactions on Systems, Man. and Cybernetics 19(3), 462–472 (1989)
5. Crandall, B., Klein, G.A., Hoffman, R.R.: Working minds: a practitioner's guide to cognitive task analysis. MIT Press, Cambridge, Mass (2006)
6. Frank, L.: Projective Methods. Thomas, Springfield, IL (1948)
7. Teglasi, H.: Essentials of TAT and Other Storytelling Techniques Assessment. John Wiley & Sons, New York (2001)
8. Bayer, H., Holtzblatt, K.: Contextual Design: Defining Customer-Centered Systems. Academic Press, London (1998)
9. Howard, J.A.: Social psychology of identities. Annual Review of Sociology 26, 367–393 (2000)
10. Epstein, S.: Integration of the cognitive and psychodynamic unconscious. American Psychologists 49, 709–724 (1994)
11. Crossley, M.: Introducing Narrative Psychology. In: Self, trauma and the construction of meaning, Open University Press, Pl (2000)
12. Berridge, K.: Pleasure, pain, desire, and dread: Hidden core processes of emotion. In: Kahneman, D., Diener, E., Schwarz, n. (eds.) Well-Being: The Foundations of Hedonic Psychology, Russell Sage Foundation, NY (1999)
13. Carroll, J., Rosson, M.: Usability Engineering. Scenario-based development of Human-Computer Interaction. Morgan Kaufmann Publishers, NY (2002)
14. Rasmussen, J.: Outlines of a hybrid model of the process plant operator. In: Sheridan, T.B. (ed.) Monitoring behaviour & supervisory control, pp. 371–383. Plenum Press, New York & London (1976)
15. Carroll, J., Robert, L., Robertson, S., Rosson, M.: Binding objects to scenarios of use. International Journal of Human-Computer Studies 41, 243–276 (1994)
16. Larsen, R., Fredrickson, B.: Measurement issues in emotion research. In: Kahneman, D., Diener, E., Schwarz, n. (eds.) Well-Being: The Foundations of Hedonic Psychology, Russell Sage Foundation, NY (1999)
17. Loewenstein, G., Schkade, D.: Wouldn't it be nice? Predicting future feelings. In: Kahneman, D., Diener, E., Schwarz, n. (eds.) Well-Being: The Foundations of Hedonic Psychology, Russell Sage Foundation, NY (1999)
18. Louie, T., Obermiller, C.: Gender stereotypes and social-desirability effects on charity donation. Psychology & Marketing 17(2), 121–136 (2000)
19. Loewenstein, G., Schkade, D.: Wouldn't it be nice? Predicting future feelings. In: Kahneman, D., Diener, E., Schwarz, n. (eds.) Well-Being: The Foundations of Hedonic Psychology, Russell Sage Foundation, NY (1999)
20. Jo, M.: Controlling social desirability bias via method factors of direct and indirect questioning in structural equation models. Psychology & Marketing 17(2), 137–148 (2000)

How Developers Anticipate User Behavior in the Design of Assistance Systems

Cordula Krinner

Technische Universität Berlin
Prospective Design of Human-Technology Interaction Group
Franklinstr. 28–29, 10587 Berlin, Germany
Cordula.Krinner@zmms.tu-berlin.de

Abstract. This paper proposes a new perspective on the old problem of function allocation. Instead of analyzing a synchronous interaction between human and computer, we suggest it could be more helpful to reconceptualize function allocation as an asynchronous division of labor between developers/designers and operators/users of human-computer systems. A study with 31 teams of developers was conducted in order to determine the effect of various forms of contact with a system and user participation on the results of a development process. The implications of lacking expected effects for future research are discussed.

Keywords: Assistance systems, allocation of functions, developers, participative design.

1 Introduction

The problem of allocating tasks in human-technology systems seems to be a never-ending story. Fitts provides one of the oldest propositions to tackle that problem, dating back to 1951 [1]. Although following this, many and apparently novel ways of dealing with this problem have been proposed by various authors, Dekker and Woods [2] claim in an article that even some of today's research and literature on function allocation is still pursuing an "Abracadabra dream of MABA-MABA methods: put your allocation problem into our method, and the solution will emerge from the other end" [2]. Instead of pursuing the silver bullet of MABA-MABA, this paper tries to shed a new light on the problem of functional allocation in human-technology systems. We propose to reconceptualize human-technology interaction as a division of labor between developers and operators of systems.

2 Theoretical Background

2.1 Function Allocation

The term function allocation is used to describe the question which functions in a human-computer or human-technology system in general should be assigned to

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 98–107, 2007.

humans, e. g. operators or users and which to machines or computer programs. Fitts wrote a report with one of the first approaches to that question [1]. It included a listing of respective abilities of humans and machines, which is referred to as "Fitts' list" or "MABA-MABA"-list, a short form for "Men Are Better At – Machines Are Better At". The question whether humans or technological devices are superior in which area was not only addressed by Fitts and other list accounts. Sheridan and Verplank developed a taxonomy of different levels of automation [3]. An enhanced version of this taxonomy taking into account stages of human action was proposed by Parasuraman, Sheridan, and Wickens [4]. Kaber and Endsley described another way to classify automation [5]. Maybe a more philanthropic approach to classification of human-technology cooperation consists in categorizing technical *assistance* to human actions instead of *automation* as a substitute for human action [6,7]. Technical assistance can be conceptualized as a subdivision of technical automation; yet the classifications for automation mentioned above do not cover all aspects that are relevant for assistance functions or systems.

In the following paragraph, a short description of Wandke's taxonomy will be given [7]. He proposes classifying assistance systems on a total of five dimensions. The central dimension is stage of human action which is supported by an assistance system. This dimension can be further subdivided into six stages that constitute one complete action cycle:

1. motivation, activation, and goal setting
2. perception
3. information integration, generating situation awareness
4. decision making, action selection, and action execution
5. action execution
6. processing feedback of action results.

A second dimension are adjustment possibilities of an assistance system, spanning from fixed systems over adaptable to adaptive systems. The third dimension describes if an assistance system initiates supporting actions, i. e. proactive assistance, or if the user has to ask for support, i. e. reactive assistance. Dimension four describes the possibilities the user has to enter data, e. g. mono-, multimodal or no explicit input. Finally, dimension five comprehends the output of the system, e. g. mono-, multimedia or implicit presentation.

Three other approaches to the problem include designing *for* the user by following principles of user-centered design [8] and designing *with* the user, i. e. applying participative design methods [9,10]. Design might also be carried out *by* the user in that predefined components of a product can be arranged in a way suitable to the individual user. These techniques might at first seem quite distinct from list and various classification approaches, but they try to ensure optimal system performance and optimal function allocation by taking into account characteristics of (future) users.

2.2 Division of Labor Between Developers and Operators

This paper wants to open a new perspective on topics of function allocation and the cooperation between human and technology. Focus of the aforementioned theoretical accounts is a synchronous interaction between human and computer/technology. This view shall be shifted towards an asynchronous division of labor between two groups of people, namely

– those anticipating and planning systems, i. e. developers, programmers or designers, and
– those finally using the implemented systems, e. g. operators or users.

The part of the project described in this paper aims at determining a profile of resources used and contributions made by developers. In collaboration with research conducted concerning operators, it will be possible to compare resources used and special contributions made by each of these two groups.

3 Experimental Paradigm and Research Question

Using the same experimental paradigm throughout all studies, it is possible to compare performances of these two groups. The paradigm consists of a *cooperative* tracking task with decision situations and takes place in a microworld [11,12,13], cf. Fig. 1. An operator's task is to provide on-line assistance and guidance to two individuals fulfilling a simple motor tracking task cooperatively. In contrast, a developer's task is to plan in advance which technical assistance functions might be useful for the cooperatively tracking subjects (to simplify this text, those two subjects will from now on be called "microworld inhabitants" as they are conceptualized as an animate part of the microworld). Implemented assistance systems will later on assist microworld inhabitants performing the

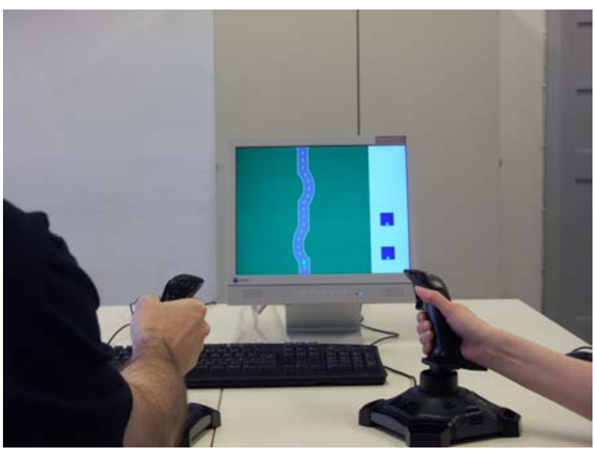


Fig. 1. The cooperative tracking task

cooperative tracking task. Thus, it is possible to compare performance under conditions of real time assistance given by an operator and technical assistance provided by a developer.

The study reported here deals with the research question how one special resource influences process and results of the development of assistance systems. The resource of interest is contact with the system and its users for whom assistance shall be created. It is hypothesized that a greater amount of contact induces more knowledge, which in turn results in better development processes and results. Thus, this study constitutes an experimental check of the tacit agreement that participative design methods [9,10] lead to systems that are more usable. This paper only reports effects of different kinds of exposure on the *results* of developing. Thus, the effects of the independent factor on one general dependent variable—kind of assistance system created—will be reported.

4 Method

4.1 Participants

62 subjects participated; as they developed assistance systems, the terms developers, subjects and participants will be used interchangeably. Subjects were advanced students of technical courses of study, e. g. electrical and mechanical engineering, information science and transportation science. Reflecting the distribution in these courses of study, 49 participants were male, 13 were female. As participants were assigned to teams of two, the relevant sample size is 31 developer teams.

4.2 Design

An experimental design with between-subjects manipulations was chosen. The manipulated factor was varied according to the three following levels:

1. Developers only read a detailed description of characteristics of the system for which they had to develop an assistance system (baseline, condition 1). This description was also given to developers in the other conditions.
2. After reading the description, they performed a cooperative tracking task themselves (condition 2).
3. They watched a screen recording of a cooperative tracking situation. In this video, typical problems of cooperative tracking were exemplified. Cooperative tracking in the screen recording was executed on a predefined route. It was identical to the route on which developers in condition two performed the tracking task on. Furthermore, teams had the opportunity to ask the microworld inhabitants questions, which were answered by the experimenter (condition 3).

Developers' task was to develop one or more concepts to assist cooperative tracking. The dependent variable was kind of assistance system developed.

4.3 Procedure

Teams were tested individually. After the arrival of both developers, they were informed about data protection guidelines as every session was completely recorded on video. Following this, the manipulation was introduced. That is, participants read only the instruction, read the instruction and performed cooperative tracking themselves for ca. eight minutes or read the instruction, watched the screen recording and asked questions. In the main part of each session, teams of developers created one or more concepts to assist a cooperative tracking task. To this end, they could use a white board or paper and multi-colored pens to take notes and draw sketches of their ideas. When teams concluded that they had completed creating their concept(s), the experimenter asked them to give a short oral summary of the system developed. This served the purpose of clarifying remaining questions of the researcher. Participants then filled in a short questionnaire on demographic variables. Finally, they received payment and were informed about background and aim of the study. Sessions lasted between one and two hours.

5 Results

5.1 Analysis of Data

Videotapes were transcribed; results of the development processes were inferred from videos, transcriptions, notes taken by participants and notes taken by the experimenter. These results were classified using the taxonomy of assistance systems mentioned before [7]. Some additional categories were inferred from the material using qualitative content analysis [14] and Grounded Theory techniques [15].

5.2 Inter-rater Reliabilities

The majority of the material was classified by one person, the author of this paper. Though, a sample (10 out of 31 teams) of the data was also analyzed by a second independent, trained rater. Inter-rater reliabilities were calculated for a total of 46 variables; Cohen's κ was used as statistical measure. κ ranged from a minimal value of .5 to a maximum value of 1. The mean κ, calculated over all variables, was .851.

5.3 General Results

First analyses were conducted in order to ensure that the three groups did not differ on any of the demographic control variables. An $\alpha = .05$ was set for all following statistical tests. No significant differences between groups could be detected concerning any demographic variable (e. g. age, gender, course of studies, number of projects participated in).

A descriptive analysis revealed the following general pattern: Some action stages seem to be "obvious" ones; these stages are supported by almost every concept and include motivation, activation, and goal setting – warning (30 out of

39 concepts support that stage) and perception (37 out of 39). On the contrary, the stage of information integration/situation awareness seems to be "invisible" as it was supported by no concept at all. Developer teams seemed more inclined to choose an assistance for the stage of decision making (26 out of 39 concepts) than not. Finally, there are stages only few developer teams decide to support. These are action execution (15 out of 39), control of effect (12 out of 39), and motivation, activation, and goal setting – activation and coaching (6 out of 39).

One hypothesis derived from the general hypothesis mentioned above was that teams with more contact with system and users might develop more than one assistance concept. The vast majority of teams created one concept, only few teams created two or three concepts [$M = 1.3,\ SD = .63$]. An analysis of variance did not show any significant difference in number of developed concepts across the three groups [$F(2, 28) = .6,\ p = .58,\ \eta^2 = .007$]. Another effect could be that more contact might result in systems that support a *broader range* of human action stages. Teams in condition two or three might have designed assistance functions that support more stages of human action. An analysis of variance was conducted in pursuing an answer to that question. Yet again, no differences between groups could be detected [$F(2, 36) = .6,\ p = .55,\ \eta^2 = .005$].

5.4 Classification of Data According to Taxonomy of Assistance

Assistance concepts developer teams had created were classified using the previously described taxonomy of assistance systems [7]. Data resulting from these ratings were exclusively nominal. Thus, statistical analyses were based on cross tabulations. As expected frequencies were smaller than 5 in more than 20% of all cases, exact tests were used. Tests were executed for each stage in the human action cycle as well as for every other dimension of the taxonomy. To maximize the amount of information used in statistical analyses, tests were executed for all concepts developed by each team. Thus, this section reports results based on $N = 39$ assistance concepts. As the pattern in results matches those of $N = 31$ concepts—each team is represented by one concept—, the bias introduced by differing numbers per experimental condition was accepted. No significant differences were found between groups concerning any stage of the action cycle, cf. Table 1. Note that no team proposed assistance concepts that could be classified as supporting processes of information integration, i. e. stage three. Thus, no analysis could be executed for that particular stage.

In summary, only one single statistically significant difference between the three experimental conditions occurred concerning the classification of solutions according to a theoretically based taxonomy of assistance. Due to the large number of significance tests conducted, this is probably a random significance. Effect sizes were constantly (very) small, only for two stages (decision making and action execution) the estimated effect sizes were medium. Apart from classifying data according to this taxonomy, other relevant aspects emerging from recorded sessions were classified using schemes derived from data. One of these will be reported to complete the results section.

Table 1. Test statistics (χ^2_{ex}), p-values and effect sizes (ES) for all characteristics of all assistance concepts, $N = 39$

Assistance for Action	Action Stage	Change-ability	Initiative	Medium	Modality	Kind of Input	Parameter
1a) Activation and Coaching	3.1	2.2	2.2	2.0	2.2	3.9	χ^2_{ex}
	.65	1.0	1.0	.94	1.0	.49	p
	.28	.24	.24	.23	.24	.32	ES
1b) Warning	.5	7.6	5.2	12.8	5.2	4.0	χ^2_{ex}
	.90	.17	.22	.51	.22	.10	p
	.11	.44	.36	.57	.36	.32	ES
2 Perception	2.6	7.2	5.8	13.7	3.4	3.4	χ^2_{ex}
	.32	.20	.16	.06	.50	.50	p
	.26	.43	.39	.59	.30	.29	ES
4 Decision Making	13.0	8.1	3.3	16.8	2.6	1.5	χ^2_{ex}
	.06	.15	.52	.047*	.71	.76	p
	.58	.46	.29	.66	.26	.20	ES
5 Action Execution	7.9	10.0	6.2	10.2	6.2	4.0	χ^2_{ex}
	.18	.05	.14	.14	.14	.10	p
	.45	.51	.40	.51	.40	.32	ES
6 Feedback	4.7	5.2	5.2	6.8	5.2	4.0	χ^2_{ex}
	.34	.22	.22	.23	.22	.10	p
	.35	.36	.36	.42	.36	.32	ES

5.5 Assistance for Cooperation

During sessions participants were often observed debating processes of collaboration, cooperation and decision making *between* the two microworld inhabitants. To consider these discussions, the concept of assistance for cooperation was introduced. It was divided into two subdivisions: One aspect was the distribution of tasks, the other aspect was the distribution of decision processes in the team of microworld inhabitants. Each of these two aspects was further subdivided according to the topics discussed by developers.

Distribution of Tasks. Two proposals were by far most popular: (1) Many developers favored an equal distribution of steering between the two microworld inhabitants; this was the default implemented in the system at the time it was presented to developers. (2) An alternative idea frequently generated by developers comprised that one microworld inhabitant should be the pilot and do most of the steering, the other one should be the copilot and act only when necessary. Statistical analyses did not show any significant differences between groups concerning this aspect [$\chi^2_{ex} = 12.9,\ p = .24,\ ES = .6$].

Decision Making in the Microworld. Decision making processes between microworld inhabitants could either not be explicitly arranged. Very few developers opted in favor of letting one of the microworld inhabitants being in command over decisions all the time. Other alternatives included that microworld

inhabitants should be informed before a decision situation arises so that they could decide what to do together; the assistance system generates an advice what to do; or microworld inhabitants take turns in deciding. This dimension was the only one in which a significant difference between groups could be detected [$\chi^2_{ex} = 16.5,\ p < .05,\ ES = .7$]. Teams who had fulfilled the cooperative tracking task themselves tended to make no suggestion how to proceed in decision situations, possibly because none of them had experienced difficulties in decision situations. Contrarily, teams in condition three preferred the solution with automatically generated advice on how to decide.

6 Discussion

To sum up the prior results section, with the exception of one variable (cooperative decision support) analyses did not yield any statistically significant differences between the three groups, effect sizes were of small or medium value at best. Thus, *results* of developing processes did not differ significantly according to the manipulation introduced, which was kind of contact with the system.

To interpret these rather devastating results, some remarks should be made. Although there were no significant differences between groups referring to each single stage of action, the general pattern described above was observed. Some action stages seem to be "obvious" ones, other appear to be "invisible". The manipulation introduced did obviously not have enough impact to change that pattern that appears consistently across all groups.

Several other reasons for the lack of expected effects can be given. First, the study reported here was a balancing act between qualitative and quantitative research. Due to restraints of time, only 10, respectively 11, teams could be assigned to each condition. It could be attributed to small size of the sample that in the rare cases of medium effect sizes no statistical significances resulted.

Second, data coded from empirical observations were exclusively categorical data, i. e. measured on a nominal level. This limits possibilities for statistical analyses, only χ^2-tests were reported here. But also more complex analyses using log linear models did not yield different results.

Third, a ceiling effect might have occurred. As the written description of the system and users' tasks in it was very detailed, even teams in the baseline condition who only received that instruction had enough knowledge to provide sufficient solutions. Thus, there was not enough room for the manipulation to result in any observable effects.

Finally, another reason for lacking effects might be adjustment/anchoring effects (cf. [16]). Prior knowledge developers brought into the situation and that was implicitly or explicitly discussed in sessions influenced generation and judgment of solutions. Examples of prior knowledge relevant for this study include current driver assistance systems in vehicles as well as PC driving simulations and racing games like "Need for speed^{TM}". These known solutions constitute an initial value/starting point for developers' judgment of their own assistance

concepts. Participants might have been unable to make sufficient adjustments, as such anchors were most likely present.

Generation and judgment processes can also be seen from the point of view of satisficing [17,18]. "Satisficing takes the shortcut of setting an adjustable aspiration level and ending the search for alternatives as soon as one is encountered that exceeds the aspiration level." ([18], p. 13). That is, if a concept created by a developer team reaches the aspiration level—known assistance systems for similar tasks—the team might have concluded its development process.

7 Concluding Remarks and Further Research Plans

Although no major differences between groups could be detected, this study will serve as a basis for two more studies. The 39 resulting assistance concepts will be evaluated in two steps. First, a qualitative expert evaluation of all concepts will be conducted. A small group of experts for creating and evaluating assistance systems will evaluate concepts with regard to how well they support a fast and accurate completion of the task and how they support cooperation and communication between participants. This study will show if the quality of assistance concepts differs and which concepts are judged to be the best ones.

Second, assistance concepts judged as the best ones will be implemented. In a within-subjects design, microworld inhabitants will track cooperatively with assistance given by technical systems and on their own. Time and error rates will be recorded to measure performance. This study will constitute a quantitative evaluation of assistance systems created by developers.

Eventually, performance of teams with assistance given by an operator can be compared to performance with assistance by a technical assistance system created by developers. Thus, a profile of contributions of each of these groups can be assembled. It can be used to show which special division of labor between developer and operator is appropriate for specific situations. If developing and designing activities proceed according to such a scheme of participation, quality of products and software can be improved. Safety for both operators and environment will be enhanced and operators might not have to fight degrading skills any more, but might be able to cooperate with developers and technologies making use of enhanced competences.

Acknowledgments. Research reported herein was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation), as a part of the program Graduiertenkollegs (Research Training Groups). Grant number: Graduiertenkolleg 1013/1 prometei.

References

1. Fitts, P.M.: Human engineering for an effective air navigation and traffic control system. Technical report, Ohio State University Research Foundation, Columbus, OH (1951)
2. Dekker, S.W.A., Woods, D.D.: MABA-MABA or abracadabra? Progress on human-automation co-ordination. Cognition, Technology & Work 4, 240–244 (2002)

3. Sheridan, T.B., Verplank, W.L.: Human and computer control of undersea teleoperators. Man-Machine Systems Laboratory report, Massachusetts Institute of Technology, Cambridge, MA (1978)
4. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics – Part. A: Systems and Humans 30, 286–297 (2000)
5. Kaber, D.B., Endsley, M.R.: The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. Theoretical Issues in Ergonomics Science 5, 113–153 (2004)
6. Timpe, K.P.: Unterstützungssysteme als interdisziplinäre Herausforderung: Einführung in die Tagung "Wohin führen Unterstützungssysteme?". In: Willumeit, H.P., Kolrep, H. (eds.) Wohin führen Unterstützungssysteme? Entscheidungshilfe und Assistenz in Mensch-Maschine-Systemen, Sinzheim, Pro Universitate, pp. 1–20 (1998)
7. Wandke, H.: Assistance in human-machine interaction: A conceptual framework and a proposal for a taxonomy. Theoretical Issues in Ergonomics Science 6, 129–155 (2005)
8. Billings, C.E.: Aviation automation: the search for a human-centered approach. Lawrence Erlbaum Associates, Mahwah, NJ (1997)
9. Falzon, P.: Enabling safety: Issues in desing and continuous design. In: Design process and human factors integration: Optimising company performance. Proceedings of the Ninth International Symposium of the ISSA Research Section, Nice, INRS (2006)
10. Muller, M.J., Wildman, D.M., White, E.A.: Taxonomy of PD practices: A brief practitioner's guide. Communications of the Association for Computing Machinery 36, 26–28 (1993)
11. Brehmer, B., Dörner, D.: Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. Computers in Human Behavior 9, 171–184 (1993)
12. Gray, W.D.: Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research. Cognitive Science Quarterly 2, 205–227 (2002)
13. Sauer, J., Wastell, D.G., Hockey, G.R.J.: A conceptual framework for designinig microworlds for complex work domains: A case study of the cabin air management system. Computers in Human Behavior 16, 45–58 (2000)
14. Mayring, P.: Qualitative Inhaltsanalyse: Grundlagen und Techniken, 4th rev. edn. Deutscher Studien Verlag, Weinheim (1993)
15. Strauss, A.L., Corbin, J.: Grounded Theory: Grundlagen qualitativer Sozialforschung. Beltz, Weinheim (1996)
16. Tversky, A., Kahneman, D.: Judgment under uncertainty: Heuristics and biases. Science 185, 1124–1131 (1974)
17. Simon, H.A.: Rational choice and the structure of environments. Psychological Review 63, 129–138 (1956)
18. Gigerenzer, G., Todd, P.M., ABC Research Group: Simple heuristics that make us smart. Oxford University Press, Oxford (1999)

Design Perspectives: Sampling User Research for Concept Development

Petri Mannonen and Mika P. Nieminen

Helsinki University of Technology, Department of Computer Science and Engineering, P.O. Box 9210, 02015 TKK, Finland
{petri.mannonen, mika.nieminen}@tkk.fi

Abstract. In user-centered design user research usually produces detailed description of the users, analysis of relevant actions and the specifics of the environment and artefacts thereof. However additional step of sampling from different viewpoints is required to more efficiently kick-start user-centered product concept development. Design perspectives, abstracted cross-category samples of user research results, are introduced as means to increase the usefulness of analyzed data during user-centered product concept development. Defining additional layer of entities in addition to conventional user research results helps to migrate from the pedant data-gathering phase to the more creative concept development phase. This paper describes definitions of entities involved in constructing design perspectives in a user research method independent manner with examples from a project developing new concepts for mobile and distributed team work.

Keywords: User research, user-centered design, concept development, data analysis, process development.

1 Introduction

User-centered design is based on the idea of developing products and services that support the user’s tasks in the context she operates in. This goal is approached by first understanding the context of use and the requirements of the users, and then making the design decisions based on that understanding. User research is the phase in user-centered design process where the knowledge of the users' tasks and environment is gathered [1].

1.1 User Research

The aim of user research is to provide sufficient understanding of the users and the context of use in order to be able to make and justify the design decisions. ISO 13407 standard describes user research as understanding and specifying the context of use. The context is defined as the characteristics of the users, tasks and the organizational and physical environment. The result of user research is typically a description document. User and organizational requirements for the product are formed based on

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 108–117, 2007.

the context of use description. However, the standard does not describe the methods used in the user research. It only suggests that the whole process should be iterative and that the description document should evolve during the whole development project [1].

Aiming for adequate understanding is the key difference between product development project's user research and traditional qualitative research, for instance anthropological studies. The main reason for this difference is the limited amount of time available for conducting the research [2]. This has lead to development of specialized research methods to gather specific type of information or to hasten the traditional qualitative research methods. For instance cultural probes focus on the personal feelings and experiences of the users [3], while contextual inquiry emphasizes the context in which work is performed [4]. The basic methodology described in the literature includes observations, interviews and questionnaires, from which a myriad of variations have sprung to tackle more specialized needs (for examples please refer to [4, 5, 6]).

User research methods descriptions usually give accurate and sometimes even practical step-by-step instructions on how to conduct the data gathering and what are the strengths of the method compared to the other user research methods. Although the research methods for gathering user and context of use information are numerous they rarely include exact instructions or suggestions how to analyze the gathered data. The authors mostly only provide general examples or rough guidelines for the data analysis. In the literature it is widely agreed that multidisciplinary processes, highly visual methods and attending to projects' special needs produce best results in user data analysis [2, 4, 6].

In qualitative research the data analysis is depicted as continuous and somewhat iterative process during which the researcher digests vast amount of information and builds up a deep understanding of the studied phenomena trough scrupulous analysis of each individual piece of data available [7]. The analysis process begins and ends with going trough the whole gathered data. At the beginning the motivation is to understand the amount of the data as well as the main themes, while at the end the results are evaluated against the original data [7, 8]. Fetterman describes the building of holistic understanding as crystallization in which the researcher condenses the information to elegant descriptions [9].

Product development is limited by availability of resources, calendar time and human resources, also during the data analysis. After the data has been gathered as efficiently as possible also the analysis must be done in rapid succession. In many cases the long analysis phase followed by a crystallization of the researchers understanding is too time-consuming for product development purposes.

Based on the themes of the ISO 13407 standard most user research results are reported with document(s) describing the user (profiles, personas, user group and role definitions, etc), environment (Environment lists, artefact lists, etc.), users' actions (task models and sequences, scenarios, etc.) and requirements derived from the preceding descriptions [4,5,6]. The usefulness of having several different representations and point of views is emphasized as in many cases it is difficult to predict exactly what sort of information is needed in the following product development phases.

Though the aim is to produce a holistic view of the users and their contexts, usually the deliverables tend to report the results as separate descriptions from predefined viewpoints (user, environment and actions).

From the viewpoint of concept development i.e. when using the user research data to create new product concepts, there seems to be a need to take the analysis a step further or at least restructure the results differently. The following chapter describes some of the special characteristics of concept development that suggest that while requirements drawn from the user research are suitable for developing predefined products, they tend to restrict the creation of new product innovations.

1.2 Characteristics of Concept Development

User-centered concept development can be described as early phase exploratory process in product development aiming to create new product innovations [10]. The motivation for user-centeredness is to introduce information about real users to the product development cycle as early as possible. Injection of user understanding to the process enables development of product ideas that respond more accurately to users' needs and comply with the context of use of the studied domain. From the company's perspective user research can also point out new product potential or new customer groups for their overall development outside the then current project. The main phases of user-centered concept development process can be outlined to include: 1) project commitment, 2) user and technology research, 3) innovation sprint, 4) concept creation and validation, and 5) project assessment. The final deliverables of this kind of process are concept descriptions detailed enough to support decision making whether to start the actual product development project or to discard the concept. [11]

From the perspective of the project team the concept development project includes three separate modes of working. During the project commitment and user and technology research one should act objectively and analytically. The idea generation phase, or innovation sprint, requires the team to be able to produce innovative and even surprising product ideas. Concept creation and validation and project assessment are a more concrete phase during which the produced ideas are evaluated and developed further and finally the whole project is wrapped up for easy adoption in the actual product development.

After user research the process' working mode changes (or more accurately is brutally rushed away) from analytical and objective research point of view to more innovative idea generation [11]. The analytical researcher's and the innovative designer's points of view are profoundly different even in situations where one and same team conducts the whole concept development process. The researcher tries to objectively understand and describe the object of research and the designer inevitability aims towards changing the world she is designing new products to.

Pre-existing know-how and roles affect on how the world and events around us are interpreted. Same subjects and events manifest themselves from different point of view based on the role of the actor (e.g. developer, evaluator, user, customer). The different points of views and frames of reference can easily lead to conflicts and misunderstandings between people. Schön claims that one way of solving these conflicts is to develop new point of view which is shared between the conflict parties or at least understood by them, hence sometimes referred as problem framing. The shared frames of reference enable the parties to work towards shared goals together [12].

In concept development projects conflicts and misunderstandings can be seen inside the concept development project in the beginning of the idea generation where in many cases there is a strong need for concrete starting points and the comprehensive user and environment descriptions seem to offer too little gripping surface for successful creative leap. The explanatory and objective user research results can make everything seem equally interesting and important.

After the innovation phase when further developing and defining the concept ideas the project team faces many decisions, which have to be based on the knowledge of users and their context of use. In decision-making and evaluation the project team has to be objective and nonbiased. Too subjective or one-sided information complicates the decision-making. The user research results have to support the objective evaluation of the concept candidates.

These above mentioned characteristics of product and concept development are in some extent noticed by user-centered product development literature. Often suggested solution to the challenges is to produce multiple models or other presentations of the findings [4]. This inevitably leads to heavy and time-consuming processes. Another solution is to start the idea generation phase (design phase) by collecting team's shared visions and using them as input for further idea generation [4]. This working method risks compromising the traceability of the process as the visions are not always based on user research findings and can embody team's prejudices.

The above concept development process sets some specific requirements for both the implementation of user research and especially for the deliverables of the research phase.

– The user research process should be as transparent as possible. Especially concept validation would gain if each individual interpretation/decision could be traced to the original observation or phenomenon.
– The deliverables should be presented in easily adoptable and understandable, preferably visual, form to minimize the time required to absorb the knowledge and to avoid false conclusions.
– The deliverables should depict the users and their tasks and contexts truthfully and comprehensively, thus enabling a solid ground for the following creative process.
– The deliverables should support the product development team while its working mode changes.
– Both the process and the deliverables should cater the overall goals of the project and the customer (i.e. maximize the utilization of earlier acquired know-how and support corporate vision).

In this paper a new abstraction, design perspective, is proposed to increase the usefulness of the findings of user research and therefore provide better starting point for concept development than current reporting practices.

Design perspectives and proposed analysis framework steers the analysis and presentations of the outcomes towards more concise concept development while balancing the user-centered approach with project and company specific development drivers in an efficient and light-weight manner.

2 Definitions

Although analyses of different qualitative data sets are different there are common features [7]. In addition of the general process of proceeding from individual observations towards more holistic view and verifying the findings with the original data also the entities evolving during the user research data analysis are similar at an abstract level.

The analysis framework depicted in figure 1 shows the main entities involved in user research data analysis from observed phenomena to categories and finally to design perspectives. The framework can be seen as hierarchical and cross-connected organization of the collected data. In the following chapters each of these entities are defined followed by a description of the process leading to design perspectives with examples.

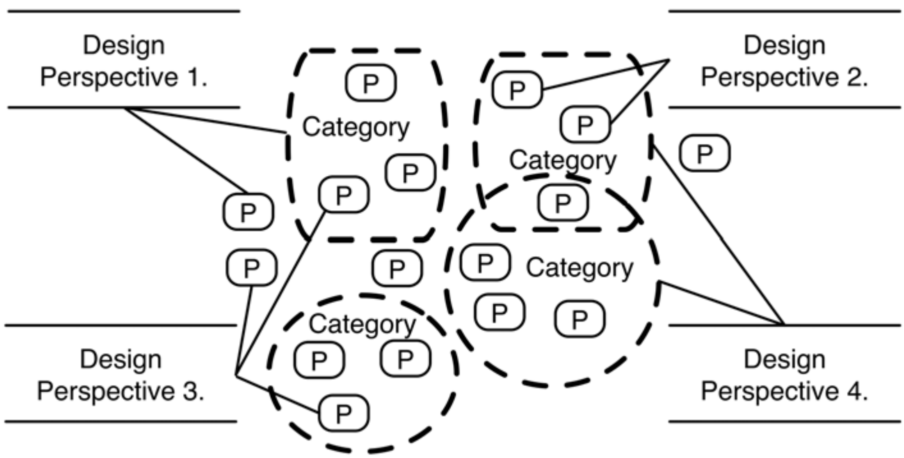


Fig. 1. User research data analysis framework

2.1 Phenomena

The phenomena (shown as ‘P’ in figure 1) are observable facts or events within the user research data. A phenomenon can be based on a single observation or be a combination of several supporting observations. Phenomena do not group the observations arbitrary together, but are more like labels for relevant observations or surfaced observation groups. However, the phenomena are not standard-sized or necessary comparable with each other. Usually the concept development project has one or more focus areas or interests. Thus the information gathered in the user research is deeper in some areas and more general in others.

2.2 Categories

During the analysis associative methods are used to form groups of phenomena hereinafter called categories. Unlike the hierarchies in affinity diagrams with all the

observations built into a single three with a top node, the categories do not need to have an existing or fabricated uniting higher level parent. Categories group together deeply linked phenomena and they may nest other categories. In figure 1 categories are marked out with dash lines.

The goals of user research (and the user-centered process) steers the analyzing process towards 'standardized' structure of user, environment and context of use related information. Additionally the projects' special goals usually show up in categorization. E.g. In a project studying distributed and mobile knowledge work the analysis of user research data produced separate categories for enablers and disablers of mobility and distributedness.

Categories and phenomena form the backbone of the analysis process leading to design perspectives. All the details about the users, environments and observed actions are included in them. Accurate documentation of the categories and phenomena enable traceability of the analysis as categories consist of phenomena and a phenomenon is directly linked to the actual observed facts or events.

2.3 Design Perspectives

Design perspectives are abstracted cross-category samples of user research results outlining discovered potential contradictions, widely visible tendencies and interesting possibilities. Design perspectives emphasize the interpretations and viewpoints of the users and the user research taskforce allowing them to contribute the tacit knowledge that could otherwise be omitted as it may fall outside the conventional reporting practices. From the concept development point of view design perspectives are a tool to kick-start the idea generation phase. They depict the user research results from different viewpoint augmenting the analyst's results with the individual user's or designer's point of view.

Creation of design perspectives is an extension of user research data analysis. Design perspectives tie together different phenomena and categories and portray the resulting combination based on the implicitly gained holistic view while roaming through the gathered user research data. Unlike insights or more developed observations noted down as a byproduct during the analysis design perspectives are formed afterwards as a separate task, thus allowing to re-examine the whole of the analyzed data and the analysis process in parallel with the thus far produced deliverables. They are created by sampling individual phenomenon or dominant groups of phenomena within a category or combining items from several categories. The chosen sampling tactics contribute to the nature of the formulated design perspectives as illustrated in figure 1:

– Design perspectives sampled from categories and singular phenomenon tend to outline potential contradictions (design perspective 1).
– Design perspectives local to a category emphasize concurring or supporting tendencies (design perspective 2) or
– When merging two or more categories even potential out-of-scope synergies (design perspective 4).

– If the sampled phenomena are local to a user or an event (design perspective 3) design perspective can emphasize an observation that would otherwise be lost during the conventional data analysis aiming for generalized consensus opinion.

It is important that the design perspectives are grounded on actual traceable phenomena or categories, thus it is possible to track down the concept candidates' underlying motives and reasoning in later stages of the overall process.

3 Building the Design Perspectives

This chapter portrays the process of building the design perspectives. The possible methods and outcomes are illustrated with examples, in italic, from authors' current project in which concepts to support mobile and distributed knowledge work are developed. The analysis framework depicted in figure 1 and definition of entities therein steer the process, but leaves fairly free hands for the implementers regarding the methods and tools used in the analysis. The discussion and suggestions included in the chapter are based on lessons learnt from current and past research projects.

3.1 Preparation of the Collected Data

Building the design perspectives begins with preparing the gathered data so that it is 'equally' and easily available for all the members of the analysis team. Strict transcripts are not always necessary, but easily understandable notes and summaries of all the materials are required. Use of highly visual representations can expedite the adoption process, for instance Graphical User Profiles or GUPs [13] have been found useful in cases where user research and concept development has been focused on individual persons. At this point the data should be kept as free as possible of interviewers/observers own interpretations, precognitions and opinions.

In our current project we used interviews and photograph probes (derived from the photography based artefact analysis [14]) as methods in user research. After the field work we produced summaries of interviews and made outlines of photograph probes debriefing interviews. The outlines showed the photographs the users' had taken and the debriefing conversation related to each photograph.

3.2 Producing Overview of the Collected Data

Next the information is shared among the analysis team. The thoroughly prepared data is handed out to every member of the analysis team for private reading and familiarizing followed by a group session where the data is walked through and explained. If the documented data is just distributed in a document format absorbing it requires a lot both from the readers and the document, as it is easy to interpret summaries and descriptions differently from its preparer's intentions.

First each researcher read the summaries and outlines by themselves and afterwards we organized a meeting in which the outlines were talked through and found misunderstandings were corrected. Special attention was paid on user's comments and citations so that the context of the citation or comment would not be lost.

3.3 Data Analysis

After developing a shared understanding of the whole gathered data the actual analysis begins. First step is to reorganize and prioritize the observations and other distinct data entities. Associative information management methods such as affinity diagrams [4] are useful and efficient way to reorganize the information. Emerging groups of related observations form phenomena. It is advisable to label phenomena (groups) with mutually understood names and explanations. The phenomena are then further organized or regrouped to categories. It is important to remember that the whole material does not need to be organized under one top category.

We used an applied version of contextual design's affinity diagrams as analysis tool. As a result of organizing and prioritizing the data, categories for enablers and disablers of mobility and distributedness emerged in addition of more traditional categories of user and environment characteristics, different work contexts and work related tasks of mobile workers.

The correctness of categories and phenomena need to be double-checked. At this point it is useful to write down the first full versions of the results (e.g. user profiles, environment descriptions, work flows, etc.). This forces to tie the open ends and check the validity of the results. The final categories and phenomena and relationships between them expose the full complexity and richness of the studied subject in a compact form for the researchers.

3.4 Creating the Design Perspectives

At this point it is advisable to take a step backwards and appreciate the effort put into gaining a full exposure to the user research data and compare the results to the objectives of the research and the project. Design perspectives allow incorporating other related information to create more accurate starting points for future product (concept) development.

Some design perspectives may surface directly from the categories as an extension of the analysis. Since the design perspectives are to outline the most interesting events and characteristic of the users and their context of use from the company's and projects' point of view there usually are a couple of candidates already present in the analysts' minds.

More illusive (and more useful) cross-category or single phenomenon design perspectives can be produced by a simple build-and-test method. Examining different categories and phenomena for supporting or contradicting aspects, selecting potentially interesting relationships to form a working hypothesis and than testing what new meanings this sample projects to the earlier gained knowledge. Observing the so far produced results from the users' point of views is another useful way to find interesting and potentially important insight and produce design perspectives.

One design perspective build in our project was a conflict between virtual working environment and real-world work context. Phenomena leading to this design perspective were:

- *Computer Supported Collaborative Work (CSCW) tools enabled collaboration between distributed workers via enabling information sharing (e.g. document databases).*

– *Team's object of work, renovating an old office building, set numerous requirements and restrictions to team members' work.*
– *Representing a physical object as complex as a building in a concrete and complete manner was difficult with the available documentation practices.*

When creating the design perspectives we realized that the centrality of the object (i.e. the building) in actual work and invisibility of it in virtual working environment produced a conflict and caused numerous problems when using the CSCW tools.

Another design perspective described the different ways the users used the same tools and the different views the users had of the tools. The different point of views and usage policies resulted in misunderstandings of the other parties' intentions.

The last phase is to document all findings, raised questions and blind spots found during the analysis process. These unknown factors need to be reconsidered later during the development and evaluation of the concepts and during the actual product development. In some cases there may even be a need to continue the research phase in order to expose the missing bits and pieces and to understand the findings more deeply.

4 Conclusions

Reporting the results of user research in a complete and easily adoptable manner is difficult, as competing viewpoints fragment the information into different documents or different section within a document. This behavior causes problems in a concept development process, as large quantities of analytical conclusions may remain as tacit knowledge inside the analysts' minds. If and when there are changes in the lineup of the development taskforce or the user research is outsourced, these valuable pieces of information are lost from the continuing development phases.

We have observed that even the most scrutinized user and context reports sometimes offer insufficient starting point for successful innovation process since the point of view from which the results are depicted is different from the point of view they are used from.

We suggest design perspectives as a simple and easily adoptable method to alleviate above problems. Design perspectives are an added abstract continuation of conventional user research allowing transition of tacit knowledge from the user research team to the group of people continuing the development process. Our experiences have shown them to better support the special requirements set by user-centered concept development processes, especially the innovation phases.

Design perspectives emphasize the tacit knowledge of the user research team and the point of views of the actual users. They also reflect the results of user research to the overall goals of a project, thus connecting separate sections of information and explicitly stating the otherwise difficult to grasp interconnection within the analyzed data.

In addition design perspectives support tracking decision making during the analysis phase by built-in traceability all the way down to an individual phenomenon. The actual phenomena and even observed events can be traced from the design perspectives. Thus mapping the design perspectives to user profiles and environment descriptions and other presentations useful in concept evaluation phase is effortless.

Acknowledgments. The authors would like to acknowledge the support from the Intact and Distributed workplace - dWork projects funded by Tekes (National technology agency of Finland) and participating companies for making this study possible. Additionally we would like to express out gratitude to Ms. Laura Turkki for her significant contributions to the initial version of the analysis framework.

References

1. ISO 13407 International Standard: Human-centred design processes for interactive systems. International Organization for Standardization, Geneve (1999)
2. Millen, D.R.: Rapid Ethnography: Time Deepening Strategies for HCI Field Research. In: Proc. DIS’00, pp. 280–286. ACM Press, New York (2000)
3. Gaver, B., Dunne, T., Pacenti, E.: Design: Cultural Probes. Interactions 6(1), 21–29 (1999)
4. Beyer, H., Holtzblatt, K.: Contextual Design: Defining Customer-Centered Systems. Academic Press, San Francisco (1998)
5. Hackos, J.T., Redish, J.C.: User and Task Analysis for Interface Design. John Wiley & Sons, New York (1998)
6. Kuniavsky, M.: Observing The User Experience: A Practitioner’s Guide to User Research. Morgan Kaufmann Publishers, San Francisco (2003)
7. Creswell, J.W.: Qualitative Inquiry and Research design: Choosing Among Five Traditions. SAGE Publications, London (1998)
8. Collier, J., Collier, M.: Visual Anthropology: Photography as Research Method. University of New Mexico Press, Albuquerque (1986)
9. Fettermann, D.M.: Ethnography Step by step. SAGE Publications, London (1989)
10. Salovaara, A., Mannonen, P.: Use of Future-Oriented Information in User-centered Product Concept Ideation. In: Costabile, M.F., Paternó, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 727–740. Springer, Heidelberg (2005)
11. Nieminen, M.P., Mannonen, P.: User-centered Product Concept Development. In: International Encyclopedia of Ergonomics and Human Factors, 2nd edn. pp. 1728–1732. Taylor & Franciss, New York (2005)
12. Schön, D.A., Rein, M.: Frame Reflection: Toward the Resolution of Intractable Policy Controversies. Basic Books, New York (1994)
13. Kankainen, T., Parkkinen, J.: GUP: graphical presentation of user profile. In: CHI ’01 extended abstracts on Human factors in computing systems, pp. 105–106. ACM Press, New York (2001)
14. Mannonen, P., Kuoppala, H., Nieminen, M.P.: Photography Based Artefact Analysis. In: Proc. INTERACT’03, pp. 833–836. IOS Press, Amsterdam (2003)

Method to Select the Most Suitable Software Tool for the Development of an Hmi Virtual Prototype

Luca Minin, Roberto Montanari, Stefano Marzani, Francesco Tesauri, and Luca Canovi

University of Modena and Reggio Emilia, Department of Science and Methods of Engineering, Via Amendola 2, Padiglione Tamburini, 42100 Reggio Emilia, Italy
{canovi.luca, marzani.stefano, minin.luca, montanari.roberto, tesauri.francesco}@unimore.it

Abstract. This paper proposes a method which allows user interfaces practitioners to choose the most suitable software tool for the development of a Human Machine Interface (HMI) virtual prototype. In the design process of a User Interface (UI), the prototyping activity represents an important way to detect design errors since the beginning of the development process, saving time and money. The choice has to take care of the technical features, designers' requirements and prototyping characteristics of the tool. The methodology includes four steps: 1) research on the most diffused virtual prototyping software employed in the prototyping activity; 2) benchmark on the technical and virtual prototyping features of each selected tool; 3) creation of a method to asses and classify each prototyping tool considering their specific characteristics; 4) application of the method to a case study.

Keywords: user centered design, virtual prototyping, user interface, HMI.

1 Introduction

In parallel with the wide spreading of the User-Centred Design approach (UCD – Norman, 1998), prototyping has become the key option to actually "user-centre" a User Interface (UI).

There are many perspectives from which the benefits of the prototyping can be outlined. From a financial point of view, for example, a long tradition of studies have revealed how cost-valuable prototyping activities are (Winters et al., 2004; Hellestrand, 1999; NASSDA, 2003); in fact, the virtual prototyping activity allows designers to correct design errors in the early stage of the process, in order to reduce the overall re-design costs.

Several prototyping tools (mainly software tools) can be exploited in all phases of the design process, from early stages to the engineering one (Dix, 2004).

To easily find design errors, it is necessary to choose the most suitable virtual prototyping software which fits not only with the requirements of the end user but

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 118–127, 2007.

also with the final employment of the virtual model. For instance, whether the virtual prototype model will be thrown away after an usability or functionality test or if it will be used as a basis for the development of a part of the final system.

A virtual prototype may be classified (Dix, 2004; Norman, 1998) according to: a) the prototyping technique used for designing the virtual model, b) the fidelity of the prototype in relation to the real system, c) the number of functions implemented by the prototype and d) the structure of the tasks.

A benchmark analysis among virtual prototyping tools was carried out in order to assess and classify them according to their technical features, designer's needs and their potential to create a virtual prototype with specific characteristics.

This classification allowed to create a method for evaluating whether a specific prototyping tool may be used for developing a prototype starting from the requirement of the users and the designers. Finally, this method was applied to a use case experiment.

2 Virtual Prototyping Tools' Sample

A benchmarking was conducted in 2005 (Minin, 2005) on academic and industrial research groups (N = 46) involved in different domains of Human Factors (HF) studies (automotive, aeronautic, nomadic devices, medical, etc.) in the US and Europe. The analysis paid particular attention to the virtual prototyping tools adopted by each group; a selection of the most widespread tools was made in order to carry out an analysis of their technical and virtual prototyping features.

The selection of the most widespread virtual prototyping tools is listed below:

- Altia Design 5.28;
- Macromedia Director Mx 2004;
- Rapid Plus 8.0;
- Gaio Protobuilder;
- MS Power Point 2003.

Subsequently, a benchmark analysis of the selected software was developed focusing on the following three macro-characteristics: (1) designer needs (2) technical features and (3) the potential of the software to create a virtual prototype of a user interface with specific characteristics.

The analysis was carried out by a group of four experts of the University of Modena and Reggio Emilia (Italy) composed by two virtual prototyping designers and two Human Factors (HF) experts.

3 Benchmark Analysis

As mentioned above, a benchmark analysis among the selected prototyping tools was carried out. Starting from previous works (Szekely, 1995; Myers, 1995; Righetti, 2006) the team of designers and HF experts selected several features of the tools

which affect the development of a prototype. These features were classified into three groups:

1. designers' needs;
2. technical features useful for virtual prototyping activity;
3. features related to the ability of the tool to create a virtual prototype with specific characteristics.

In the following paragraphs a description of the three different groups of features among the prototyping tools is depicted. In order to compare the prototyping tools, firstly a method for the evaluation of the features is presented.

Finally, the tables regarding the benchmark analysis are presented.

3.1 Method for the Evaluation of the Features

Each feature which belongs to the first and second group above mentioned is described by only one of the following characteristics:

- *Availability/Accessibility*: features could be totally available, partially available or absent in the prototyping tool. For instance, not all the tools allow to create automatic data logs which record the interaction between the prototype and the final user;
- *Complexity*: for available/accessible features is specified whether it is very complex, on the average complex or simple to be used. For instance, some prototyping tools allow to program the logic of the prototype only through a low-level programming language (complex to be used).

The degree of Availability/Accessibility and Complexity defined for each feature are based on a 3-level color scale, shown in the following table.

Table 1. Three level color scale

Colour	Availability/Accessibility	Complexity
Green	The feature is totally available in the prototyping tool.	The complexity of use of the feature is low.
Yellow	The feature is partially available in the prototyping tool.	The complexity of use of the feature is on the average.
Red	The feature is not available in the prototyping tool.	The complexity of use of the feature is high.

3.2 Designers' Needs

In the following table, the most relevant designers' needs and the related characteristics (in brackets) are numbered and described.

Table 2. Designers' needs

Number	Description of the feature
0	*Computer knowledge* (complexity). Knowledge needed for the prototype development: graphic editing (green), graphic and high-level programming (yellow) or graphic and low-level programming (red).
1	*Programming language* (complexity). The logics of the prototype can be developed: only through a low-level programming language (red), high level (green) or the tool offers both programming levels (yellow).
2	*Complexity and accuracy of the tutorial* (complexity). The use of the tutorial is: complex (red), on the average complex (yellow) or easy (green).
3	*Availability of a tool for development of the graphic interface* (Availability/Accessibility). A tool for the development of all the prototype's graphic interface is: available inside the prototyping software (green), only simple graphics can be created (yellow) or all the graphics have to be created with an external tool (red).

3.3 Technical Features Useful for Virtual Prototyping Activity

In the following table, the most relevant technical features and the related characteristics (in brackets) are numbered and described.

Table 3. Technical features

Number	Description of the feature
4	*Generation of a stand-alone application* (Availability/Accessibility). The tool allows to create a stand alone application of the prototype which can be loaded in a computer: without licence (green), only with licence (yellow) or a stand alone application can not be created (red).
5	*Generation of programming code* (Availability/Accessibility).The tool allows to generate code of the prototype: in different programming languages (green), in only one language (yellow) or no code generation is possible (red).
6	*Connection to external applications* (Availability/Accessibility). The tool allows to connect the prototype to other applications able to manage: both the graphic and the logic through specific API (green), either of the two (yellow) or the connection is not available (red).
7	*Library of graphical objects with already implemented programming code* (Availability/Accessibility). This kind of library is available and the programming code of the objects is editable (green), not editable (yellow) or the library is not available (red).
8	*Data log for usability and functionality tests* (Availability/Accessibility). An automatic log of the interaction between the user and the prototype is provided (green), only a manual log is provided – that is, the log has to be configured using programming languages (yellow), or it is not possible to generate a data log (red).

Table 3. (*continued*)

9	*Generation of specification documents* (Availability/Accessibility). Specification documents related to the functionality and the components of the prototype with an high detail can be generated (green), can be generated with a low detail – that is, only the components but not the functionality (yellow) or specification documents can not be generated (red).

3.4 Features Related to the Potential of the Software to Create a Virtual Prototype for a Specific Kind of Use Employment

In this section, the features of the tools which allow the designer to create prototypes with specific characteristics are described. These characteristics, selected from the literature (Mayhew, 1992; Sommerville, 1995; Dix, 2004; Norman, 1998) are the following:

A. the prototyping technique (Throw-away, Incremental or Evolutive);
B. the fidelity (High, Medium or Low);
C. horizontal or vertical prototypes;
D. the structure of the tasks.

In the following paragraph each point of the list is described. The team of designers and HF experts defined a correlation between specific designers' needs and technical features with the prototyping characteristics listed above. In the following paragraph, each feature described in Table2 and Table3. Table4 call back the related number (e.g. feature 4, feature 6, etc.).

3.4.1 Prototyping Technique

The objective of the *throw-away prototyping* is to understand the system requirements. Starting from a stand alone application of the prototype, iterative tests with the end users on the usability and functionality of the final system are carried out. Finally, the specifications of the final system are defined. Thus, the prototyping tool should provide features 4 and 8 for the testing activity and feature 9 for the development of the system's specifications.

Incremental prototyping aims at developing a virtual prototype for each part of a modular system (i.e. a system composed by several components). Starting from an overall architecture of the final system, each prototype is created and tested with users (features 4 and 8 required); then it is delivered in increments with the related specification documents (feature 9) and the programming code (feature 5) is used for the development of a part of the final system.

The objective of *evolutionary prototyping* is to deliver a working system to end-users. The technique is similar to the incremental but with an evolving design of the overall system. Then, the same features of the incremental prototyping should be provided. However, the specification documents and the generated programming code are related to the overall system and are delivered at the end of the iterative design activity.

The following table is a summary of the technical features required for the development of the three prototyping techniques. The numbers of the columns

represent the technical features listed in Table2, Table3 and Table4. As may be seen in Table1, “G” and “Y” correspond to Green and Yellow, the colours which identify that a specific technical feature is available in the prototyping tool. An empty cell means that the technique does not require the specific technical feature.

Table 4. Prototyping techniques (rows) and required technical features (columns)

	4	**5**	**8**	**9**
Throw away	G,Y		G,Y	G,Y
Incremental	G,Y	G,Y	G,Y	G,Y
Evolutive	G,Y	G,Y	G,Y	G,Y

3.4.2 Fidelity of the Prototype

Regarding the fidelity of the prototype, for each level (High, Medium, Low), the prototyping tool should provide the following features:

- *High fidelity*: the prototype should be similar to the final system and it should work in the same way. Feature 4, 5, 6 are required. The tool should have all the features required to develop a prototype which can also be used as a basis of the final system.
- *Medium Fidelity*: useful for testing design concepts early in the design process. Features 4, 6 are required. The prototype should look like the final system, but it is not required it will also be the basis for the development of a part of it, then the prototype does not provide all the interactions and functions of the final system.
- *Low fidelity*: it represents a mock-up of a system (or part of a it) used to support user studies. Feature 4 is required. The prototype should be developed in a rapid way. It represents only a set of functions or interactions of the final system.

As for Table 4 Table 5, the technical features of the prototyping tool required for the development of a High, Medium or Low-Fi prototype are depicted below.

Table 5. Level of fidelity (rows) and required technical features (columns)

	4	**5**	**6**
High-Fi	G,Y	G,Y	G,Y
Medium-Fi	G,Y		G,Y
Low-Fi	G,Y		

3.4.3 Horizontal/ Vertical Virtual Prototypes

An horizontal prototype displays “breadth” of functionality, that is, broad but only top-level functions of the final system. In a vertical prototype full functionality and performance of a small part of the final system are implemented. Starting from this definition, the ability of a prototyping tool to allow the development of prototypes with these characteristics was estimated. For instance, Altia Design has a set of applications which allows to easily develop multilayered virtual prototypes although the levels may only be displayed and managed by the designer one by one and additionally, the workspace is quite limited in width (i.e. it is difficult to manage a wide range on functions characterized by different level of deepness). Hence, this tool is a good solution for the development of a deep vertical prototype but not an horizontal one.

3.4.4 Structure of the Tasks

When the user has to complete a specific task, he/she has to perform a sequence of interactions with or without the help coming from the system (e.g. affordances and feedbacks). Alternatively, a part of the task can be automated (less interactions). The more the task is automated, the easier the structure of the task will be. The ability of the tool to allow the designer to develop prototypes with an easy structure of the tasks is evaluated. If the level of the programming language for the development of the prototype’s logic is medium-high and a library of graphical objects with already implemented programming code is available in the tool (point 1 and 7 of Table2, Table3), then a more simple structure of the task can be implemented easily since some tasks may be automated rapidly.

Table 6. Structure of the tasks (rows) and technical features/designers' needs (columns)

	1	**7**
Easy structure of the task	G,Y	G,Y

3.5 Benchmark Among the Prototyping Tools

Finally, all the above mentioned features were analyzed among the selected tools in order to assess the overall characteristics. Three tables for the benchmark analysis were prepared:

- the first including the comparison of the features related to the designers’ needs Fig.1;
- the second including the technical features Fig.2;
- the third including the features related to the ability to create prototypes with specific characteristics Fig.3.

The last table is the result of the combination of the data contained in the first two tables, as described in the previous paragraph. A “X” symbol in the table means that the feature is available, otherwise the cell is empty (not available).

A part of the three tables is depicted in the following figures; in the original tables all the features of the previous paragraph were reported with a small description.

	Altia Design	Macromedia Director Mx 2004	Rapid Plus 8.0	Protobuilder	MS Power Point 2003
Computer knowledge	R	Y	R	Y	G
Programming language	Y	G	Y	G	G
Complexity and accuracy of the tutorial	G	Y	G	G	Y

Fig. 1. Benchmark on the designers' needs

	Altia Design	Macromedia Director Mx 2004	Rapid Plus 8.0	Protobuilder	MS Power Point 2003
Generation of a stand-alone application	G	G	G	Y	R
Generation of programming code	G	Y	G	Y	R
Connection to external applications	G	Y	Y	R	R

Fig. 2. Benchmark on the technical features

	Altia Design	Macromedia Director Mx 2004	Rapid Plus 8.0	Protobuilder	MS Power Point 2003
Throw away	X	X	X	X	X
Incremental	X	X	X	X	
Evolutive	X	X	X	X	
High fidelity	X				
Medium fidelity	X	X	X	X	

Fig. 3. Benchmark on the prototyping characteristics

4 Method to Asses and Classify Each Prototyping Tool Considering Their Specific Characteristics

The method to select the best prototyping tool for the development of a HMI prototype follows the steps:

1. definition of the characteristics of the prototype (see § 3.4);
2. definition of the skills and the requirements of the team which design the prototype (see § 3.2);

3. definition of the technical features required for the creation of the prototype (see § 3.3);
4. selection of the prototyping tool which fits with the characteristics defined in steps 1 to 3 (see § 3.5).

Each step of the list is described in the previous paragraphs. At step 4, the method was applied to a selection of prototyping tools but it can also be enlarged to other software. An application of the method is described in the following paragraph.

4.1 Use Case Application

The above mentioned method was applied to the development of a virtual prototype of an in-vehicle display.

The designers and the project owner required that a High-Fi and vertical prototype had to be created through an Evolutive technique. The team of designers was composed by three computer scientists and one HF expert. It was required to develop at first a stand alone application of the system for usability tests. Then, after a re-design of the system based on the results of the test, the owner required to connect the prototype to the vehicle network and test the functionalities of the prototype in the real use environment.

By comparing these requirements with the features of the prototyping tools depicted in the figure of § 3.5, the selected tool was Altia Design. In fact, it would allow to create virtual prototypes with the required characteristics. Moreover, it is possible to connect Altia to other software, in particular Matlab Simulink, and to create an interface with the vehicle network. After that, a stand alone application and data-log for usability tests can be provided. Finally, using Matlab Simulink and Deepscreen (i.e. the code generator tool of Altia Design), the programming code of the prototype can be generated and used as a basis for the development of the real in-vehicle interface.

5 Conclusions

This paper proposes a solution to help designers in choosing the most fitting software tool for the development of a virtual prototype with specific characteristics. Analyzing the technical features and the designers' requirements with certain degrees of availability, accessibility and/or complexity, it is possible to assess whether a prototyping tool is able to create a horizontal/vertical prototype with a high/medium/low-fidelity or an easy structure of the task. Moreover, it is possible to define which prototyping techniques the specific tool allows to use.

Since there are a lot of different prototyping tools with different characteristics, it is important to find the most suitable solution which fits with the expectations of the designer.

References

1. Norman, D.: The Design of everyday things. MIT press, Cambridge, MA (1998)
2. Winters, F., Mielenze, C., Hellestrand, G.: Design Process Changes Enabling Rapid Development. In: Proc. Convergence 2004, pp. 613–624. Society of Automotive Engineers, Warrendale, PA (2004)

3. Hellestrand, G.: The Revolution in Systems Engineering. IEEE Spectrum, 43–51 (September 1999)
4. NASSDA, Maximizing Silicon ROI: The Cost of Failure and Success, Nassda White Paper WP020522-1A (2003)
5. Dix, A., Finlay, J., Abowd, G., Beal, R.: Human Machine Interaction. MacGrow-Hill, New York (2004)
6. Minin, L.: Virtual prototyping in the human-machine interaction: a benchmark analysis and tests for the selection of the virtual prototyping tool which fits with the requirements of a user interface, M.S. Thesis, University of Modena and Reggio Emilia, Italy (2005)
7. Szekely, P.: User interface prototyping: Tools and techniques (1995)
8. Myers, B.: User interface software tools. ACM Transactions on Computer-Human Interaction 2(1), 64–108 (1995)
9. Righetti, X.: Study of Prototyping Tools for User Interface Design, Thesis, University of Geneve (2006)
10. Mayhew, D.: Principles and guidelines in Software User Interface Design. Prentice Hall, Englewood Cliffs (1992)
11. Sommerville: Software Engineering, 5th edn. (1995)

Intuitive Use of User Interfaces: Defining a Vague Concept

Anja Naumann[1], Jörn Hurtienne[2], Johann Habakuk Israel[2], Carsten Mohs[2], Martin Christof Kindsmüller[3], Herbert A. Meyer[4], Steffi Hußlein[5], and the IUUI Research Group

[1] Deutsche Telekom Laboratories, TU Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany
anja.naumann@telekom.de
[2] Center of Human-Machine Systems, TU Berlin, Franklinstr. 28-29, 10587 Berlin
{joern.hurtienne, johann.habakuk.israel, carsten.mohs}@zmms.tu-berlin.de
[3] artop GmbH, HU Berlin, Christburger Str. 4, 10405 Berlin, Germany
mck@imis.uni-luebeck.de
[4] Institute for Multimedia and Interactive Systems, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany
meyer@artop.de
[5] Interaction Design Lab, University of Applied Sciences Potsdam, Pappelallee 8-9, 14406 Potsdam, Germany
husslein@fh-potsdam.de

Abstract. In this paper we present a general definition of the concept ‘intuitive use of user interfaces' on the basis of our current interdisciplinary work. ‘Intuitive use’ is regarded as a characteristic of human-machine systems. It refers to a special kind of interaction process between users and technical systems that use the users' intuition. The main part of the paper deals with central aspects of this definition in detail and discusses pre-conditions and restrictions of the use of the concept. The main aspects that we discuss are the design of technical systems, application and non-conscious use of previous knowledge, intuition as a non-conscious process, interaction, and effectiveness. We complement this discussion by addressing the relationship between aesthetics and intuitive use.

Keywords: aesthetics, effectiveness, human-machine interaction, intuition, knowledge, non-conscious, usability, user interfaces.

1 Introduction

With the increasing number of functions and uses of different technical systems in daily life, the concepts of 'intuitive' and 'intuitive use' are used more and more often as an assessment criterion for these technical systems and for User Interface (UI) requirements. Frequently, these concepts are attributed to an interface itself,

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 128–136, 2007.

e.g. 'Firefox 1.5 has an intuitive interface' (www.mozilla-europe.org, 2006), without specifying precise aspects of use or user groups. If one asks users of interactive systems or usability experts what they think about 'intuitive use', one most likely gets answers like: 'acting by intuition', 'acting on instinct', 'using without guidance/explanation', 'acting as a matter of routine', or 'using automatically' [1]. All of these statements suggest that 'intuitive use' should not demand high cognitive resources (anymore), but rather work on a skill based or maybe rule based level [2].

In the scientific literature however the concepts of 'intuitive' and 'intuitive use' are commonly avoided by researchers because they are regarded as vague, ill defined or non-scientific. Facing this situation the IUUI research group (Intuitive Use of User Interfaces) has made it its business to explore the usefulness of the term 'intuitive use' as a well defined scientific concept.

We started our interdisciplinary endeavor as a group of psychologists, computer scientists, engineers, and designers without knowing whether we ultimately would be able to come up with a common concept of the term 'intuitive use' which was shared by the whole group, or whether we would convince ourselves to avoid this term in the future, or (even worse than that) whether we would neither be able to agree on the usefulness of the concept nor be able to avoid it completely. After almost one year of work, we are now not only presenting the second refinement of our definition of 'intuitive use', but we are also confident that it is possible to further develop this concept to create guidelines for the design of intuitively usable systems and devices.

In the following, we start by presenting our current definition of the intuitive use of user interfaces. In the main part, we explain and discuss central aspects of this definition in detail. These main aspects are: the design of technical systems, the application of previous knowledge, intuition as a non-conscious process, interaction and effectiveness. Though we do not regard aesthetics as an aspect of intuitive use, it is obvious that, from a design perspective, the relationship between aesthetics and intuitive use has to be clarified. We address this issue in a separate section. The paper is rounded off with a summary and an outlook of what has to be done.

2 Intuitive Use of User Interfaces: Definition Version 1.1

‘A technical system is, in the context of a certain task, intuitively usable while the particular user is able to interact effectively, not-consciously using previous knowledge.’

The basis of our definition is the conclusion that only human information processes can be labeled as intuitive. The term 'intuitivity', that has been used more frequently in recent years as a product feature, should thus be avoided. 'Intuitive use' can only be used in the context of task, user, environment or technical system. More precisely, intuitive use can only be attributed to the human-machine interaction in a certain context, for the achievement of objectives, but not to a technical system per se.

2.1 Design of Technical Systems

Intuitive usability can only be attributed to the human-machine interaction in a certain context. Nevertheless, there are some aspects which should be considered when designing a technical system.

The design space for technical systems can be described in terms of the layers given by Foley and van Dam [3] and modified by Buxton [4] in their human-computer-interaction model: *conceptual*, *semantic*, *syntactic, lexical,* and *pragmatic (physical)* layers. Each layer has its own characteristics in respect of intuitive interaction. The *conceptual* layer describes the main concepts (mental model of the users) of the interactive system, e.g. spreadsheet applications, text editors, graphical tools. The *semantic* layer defines the functionality of the system, sequences of user actions and system responses. The *syntactic* level defines interaction tokens (words) and how to use them to create semantics. The *lexical* layer describes the structure of these tokens and the *pragmatic* layer [4] describes how to generate them physically by means of user actions and I/O elements, e.g. displays and input devices.

Designers of technical systems should carefully go through these layers, starting from the conceptual layer, and carefully provide solutions for each layer and the mappings between them [5]. This process helps to prevent 'apples and oranges' types of solutions and to come up with applicable catalogues of interface solutions. Also the type of mapping between the layers is crucial for intuitive interaction. Modifications made at one level often have strong effects on other layers, e.g. changing the input device may strongly affect the interaction syntax [4].

There are hardly any guidelines for the development of intuitively usable interfaces. Blackler, Popovic and Mahar [6] developed three principles for applying intuitive interaction to interface design: 1. Use familiar symbols (lexical), 2. Make obvious what less well-known functions will do (semantic), and 3. Increase consistency between different parts of the design (lexical, syntactic). They found that the appearance (shape, size and labeling) of features most affects intuitive use (pragmatic, lexical layer). Based on a survey among HCI experts, Mohs et al. [1] list seven criteria which influence intuitive interaction. Five of them are (sub-) criteria of the ISO 9241-110 [7] usability criteria: suitability for the task, conformity with user expectations, and self-descriptiveness; others are related to affordances and Gestalt laws. Hornecker and Buur [8] refer to the vertical design mapping, the *perceived coupling* between the physical (here: tangible) and the semantic layer. Hurtienne and Israel [9] suggest the application of image schemas (pragmatic, lexical, syntactic) and their metaphorical extensions (semantic) for the design of intuitively usable interfaces. We still see a high demand for comprehensive catalogues and guidelines in this field.

Our definition of the intuitive use of technical systems covers primarily the *interaction problem* [10] defined by the pragmatic, lexical, syntactic and semantic layer. We assume that intuitively usable interfaces free more cognitive resources for the solution of the *overall problem* [10] in the conceptual layer.

2.2 Continuum of Knowledge

Users can interact with a technical system effectively and intuitively when applying their *previous knowledge* to a certain situation. This previous knowledge may stem

from different sources. These knowledge sources can be classified along a continuum from *innate* knowledge, knowledge from embodied interaction with the physical world (*sensorimotor*), and *culture* to professional areas of *expertise*. On each of the last three levels there might be specialist knowledge about using the corresponding *tools* and technologies (see Figure 1).

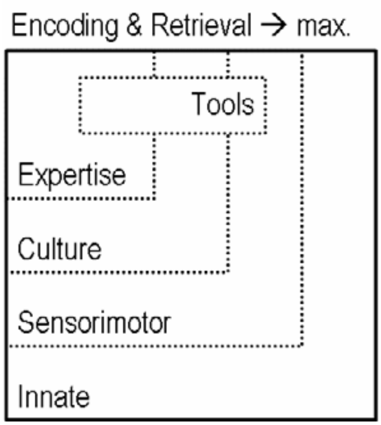


Fig. 1. Continuum of knowledge in intuitive interaction

The first and lowest level of the continuum consists of *innate* knowledge - ‘acquired’ through the activation of genes or during the prenatal stage of development. Generally, this is what reflexes or instinctive behaviour draw upon. Purists will see this as the only valid level of knowledge when talking about intuitive interaction, because it assures universal applicability and non-conscious processing. The next level is *sensorimotor*. It consists of general knowledge, which is acquired very early in childhood, and is from then on used continuously through interaction with the world. Children learn for example to differentiate faces; they learn about gravitation; they build up concepts for speed and animation. Scientific notions like affordances [11] and image schemata [12] reside at this level of knowledge. The next level is about knowledge specific to the *culture* an individual lives in. What is known within the western group of cultures is not necessarily equivalent to the knowledge of people in eastern cultures (e.g. the appropriate colour at funerals). The most specific level of knowledge is *expertise,* that is specialist knowledge acquired in ones profession, for example as a doctor, mechanic, or accounting clerk; and in hobbies (e.g. riding, surfing, online-gaming). Across the sensorimotor, culture and expertise levels of knowledge, we also distinguish knowledge about *tools.* Tool knowledge seems to be an important reference when designing user interfaces. At the sensorimotor level there are primitive tools like sticks to extend one’s reach and stones to be used as weights. At the culture level, we find tools commonly used by people, like ball point pens for writing, pocket lamps for lighting, or cell phones for communication. At the last stage, there is the knowledge acquired from using *tools* in one’s area of expertise, for example image editing tools, enterprise resource planning (ERP) systems, or CNC machines. Even within the same domain of expertise

(e.g. graphic design) there may be differing knowledge on the tool level of the continuum, depending on the kind of tools used (e.g. Corel Paint Shop vs. Adobe Photoshop).

The continuum of knowledge has an inherent dimensionality. The frequency of encoding and retrieval of knowledge increases from the top to the bottom of the continuum. Then, the further we rise towards the top level of the continuum, the higher the degree of specialization of knowledge and the smaller the potential number of users possessing this knowledge. But still, on each level of the knowledge continuum we may assign ‘intuitive use’ according to the above definition – as long as it is *non-consciously* applied by users.

2.3 Using Previous Knowledge Non-consciously

The application of knowledge by the user within the interaction with the technical system may be non-conscious from the beginning (as with reflexes) or may have become non-conscious due to frequent exposure and reaction to stimuli in the environment: the more frequent the encoding and retrieval was in the past, the more likely it is that memorised knowledge is applied without awareness by the user. Knowledge at the expertise level is acquired relatively late in life and is (over the life span) not as frequently used as knowledge from the cultural or sensorimotor level. As learning theory suggests, knowledge from the lower levels of the continuum is therefore more likely to be applied non-consciously than knowledge from the upper levels. If the non-unconscious application of knowledge is a precondition for intuitive use, it will be more common to see intuitive interaction involving knowledge at the lower levels of the continuum.

Limiting ‘intuitive interaction’ to the lower levels of the knowledge continuum which are still consciously aware does have further advantages:

- The further down we move on the continuum, the larger and more heterogeneous the user groups we can reach. While almost everyone will have an understanding of ‘verticality’ (sensorimotor level), not everyone understands the Corel Paint Shop software package (tool/expertise level).
- Instead of being required to analyse the previous knowledge and experience of the specific target user group, designers might simply refer to rules generated from findings about the general structure of human knowledge (i.e. general human knowledge on the sensorimotor level).
- Extremely frequent encoding and retrieval events lead to a higher robustness of information processing and automatic processes. In situations of high mental workload and stress, a fall-back on lower stages of the knowledge continuum will occur. This will be especially important to the design of systems with a high requirement for security (control of aircraft or of nuclear power plant).
- Non-conscious processing of user interface elements in general means less workload on the cognitive processing capacity. Thus, more cognitive resources will be available for solving the working task at hand, instead of wasting time and mental effort on figuring out how a piece of technology works.

2.4 Interaction

In the context of intuitive use of a technical system, we understand interaction as a two-way exchange of energy and information between human and product. From the human perspective, this exchange is target-oriented.

Looking at the human aspect, the following can be stated: Tasks or activities have a hierarchical structure [13]. This is shown in Figure 2. Tasks consist of an entirety of actions which are subordinated to sub-goals. These sub-goals can be derived from a superordinate conjoint goal. The role of this conjoint goal can be taken up by the motive of the task. Actions, in turn, are self-contained entities of the task, consisting of sub-actions or operations. These operations are dependent on the task, because their results are not consciously anticipated as a goal. Also, they are inherently regulated by triggering mechanisms (e.g., 'Traffic light red – brake'). Thereby, a participation of transient sub-goals is possible. Operations include several motions or single mental processes, for example conclusions, at a time [13].

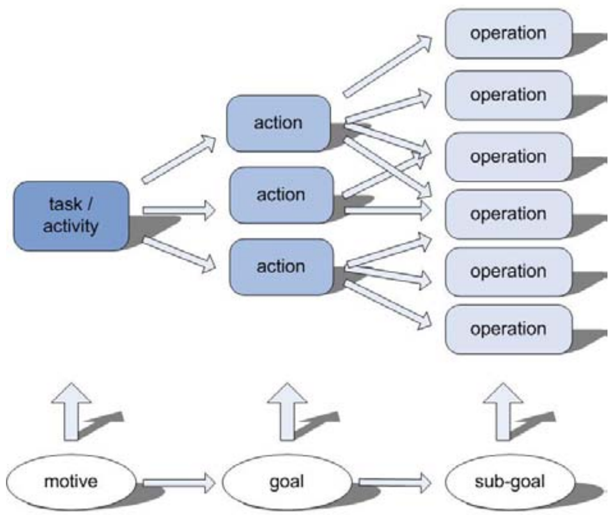


Fig. 2. Hierarchical structure of the task in the Action Regulation Theory [14] adopted from Walliser [15]

In our understanding, intuitive usability of a technical system, in the first instance, refers only to the level of operations as a part of more complex actions. On the level of operations, we think, intuitive use can be made measurable in the simplest way.

2.5 Effectiveness

The rating or assessment of the procedure of intuitive interaction can be done by its effectiveness. According to the ISO standard series 9241 the term *effectiveness* is mentioned as "the accuracy and completeness of users' tasks while using a system" (DIN EN ISO 9241-11 [16]). In this sense, it is only possible to call an interaction

intuitive if it leads the user to adequate, exact and complete interaction results. However this understanding doesn't necessarily mean that an intuitive procedure corresponds to something that system developers would call 'the optimal way'.

Furthermore, we assume that users experience intuitive interactions as *efficient.* Whereas it is important to point out the multi-dimensional character of the concept of *efficiency* and especially the orthogonality of dimensions like time, cognitive resources, material resources, energetic resources and financial resources. Under that premise, it is possible to vary each dimension extensively, independently from the other. Therefore an operation in human machine systems is meant to be intuitive if the demands of cognitive resources are minimal even if it causes a higher investment of other dimensions like e.g. time.

3 Viewpoint: Aesthetics and Intuitive Use

Though we do not regard aesthetics as an aspect of intuitive use and therefore do not include it in our definition, we also want to look at the relation between aesthetics and intuitive use from the design perspective. Do the aesthetic design characteristics of interactive products have an effect on the intuitive use of products?

Aesthetics is defined as the branch of philosophy "that in its broadest sense, deals with the general questions of art […], and that in its narrow sense, with how things are known via sensory perception […]." [17]

Aesthetic parameters include form, audio, smell, color, proportion, size, material and surface qualities. The appearance, composition and characteristics of an application or product are influenced by many combinations of these design principles and developed, together with the information architecture and the technical solutions, in an iterative product development process. Or put more simply, aesthetics concerns all parameters of interaction and appearance.

The design of intuitively usable interactive information systems and devices takes into consideration all the human senses involved, as with, for example, the visual representation of a communication process. In each sensory system an aesthetic judgment is made. Aesthetic design directly addresses human sensory perception and triggers sensations and immediate evaluations of the user experience. The higher the compatibility of the offered interaction with the user's personal and cultural habits, experiences and emotional state, the more intuitive will be the usage.

It thus follows that the aesthetics of an interactive offering is an essential influencing parameter on the degree of intuitive use. If the user does not perceive objects and signs as attractive and usable, or at least familiar, then the application or product has almost no chance of being used intuitively by the user or indeed of triggering a positive user-experience. But just what role does aesthetics play in the discussion on intuitive use? If one assumes – as we do - that the degree of intuitive use is dependent on how much previous knowledge can be implicitly recalled in the interaction process, then aesthetics should be considered to be even more important.

Aesthetic qualities trigger previous knowledge as aesthetic judgments are at the start of individual perception processes: When a user begins an unknown process or encounters an unknown device, he or she looks at it, maybe touches it, listens to it, or even smells it. Without these sensory steps, that ask questions of the device and which

are answered by its aesthetic qualities, the user cannot get close to the technical system. If none of the impressions of the device are compatible with the user's previous experience, he or she will not be able to intuitively get close to the device and its significance - as a result the user will have to refer to the online help or manual. The user feels subjectively represented by an aesthetically designed product that triggers positive memories and associations. Aesthetics is the key to the technology that lies behind.

4 Summary and Outlook

In our definition of 'intuitive use' we state that this term can only be attributed to the human-machine interaction in a certain context, for the achievement of objectives, but not to a technical system per se.

In the present paper, we explained central aspects of the definition in detail. So, we assume that intuitively usable interfaces free more cognitive resources for the overall problem, as part of the interaction problem in human-computer interaction [10]. Also, we identified several relevant levels of origin of previous knowledge used in intuitive interaction, whereas knowledge from the lower levels of the continuum is more likely to be applied non-consciously. Regarding the interaction itself, intuitive usability, in the first instance, refers only to the level of operations as a part of more complex actions in the Action Regulation Theory [13]. Furthermore, we refer to interaction as intuitive if it leads to sufficiently accurate and complete interactions for the user. Concerning efficiency, an interaction option is intuitive when users' cognitive load is reduced.

We also think there is a relation between aesthetics and intuitive use. Thus, we took a first step towards the clarification of this relation from the design perspective.

The definition and its explanation display the current state of work of our IUUI research group. The content discussion within the research group and with external researchers interested in the topic is continuing and will result in a modified version of the definition (version 2.0).

Based on the development of a concrete definition of 'intuitive use', it will be possible to formulate design principles for intuitive human-machine interaction. Accordingly, in a next step we will analyze which existing usability concepts support intuitive use in terms of our definition. Therefore, we will focus more on the central usability specifications in the ISO standard (DIN EN ISO 9241-11 [16]): effectiveness, efficiency and user satisfaction. In addition, we will look more closely at the relationship between aesthetics and intuitive use.

Acknowledgments. We would like to thank all members of the IUUI Research Group (www.iuui.de) for their contribution to the vivid discussion on intuitive use and intuitivity.

References

1. Mohs, C., Hurtienne, J., Scholz, D., Rötting, M.: Intuitivität - definierbar, beeinflussbar, überprüfbar [Intuitivity – definable, influenceable, checkable]. In: Useware 2006 - VDI Berichte Nr. 1946, pp. 215–224. VDI-Verlag, Düsseldorf (2006)
2. Rasmussen, J.: Information processing and human-machine interaction: An approach to cognitive engineering. North-Holland, Amsterdam, NL (1986)

3. Foley, J.D., van Dam, A.: Fundamentals of Computer Graphics. Addison-Wesley, Reading, MA (1982)
4. Buxton, W.: Lexical and Pragmatic Considerations of Input Structures. Computer Graphics 17(1), 31–37 (1983)
5. Shneiderman, B.: Designing the user interface. Strategies for effective human-computer interaction. Addison-Wesley, Reading, Mass (1998)
6. Blackler, A.L., Popovic, V., Mahar, D.P.: Intuitive Interaction Applied to Interface Design. In: Proc. International Design Congress - IASDR 2005, Douliou, Taiwan (2005)
7. DIN EN ISO 9241-110. Ergonomics of human-system interaction – part 110: Dialogue principles. German version. Beuth, Berlin (2006)
8. Hornecker, E., Buur, J.: Getting a Grip on Tangible Interaction: A Framework on Physical Space and Social Interaction. In: Proc. CHI 2006, pp. 437–446. ACM Press, New York (2006)
9. Hurtienne, J., Israel, J.H.: Image Schemas and Their Metaphorical Extensions – Intuitive Patterns for Tangible Interaction. In: Proc. Tangible and Embedded Interaction 2007, ACM Press, New York (2007)
10. Streitz, N.A.: Cognitive ergonomics: An approach for the design of user-oriented interactive systems. In: Klix, F., Wandke, H. (eds.) MACINTER I, pp. 21–33. North-Holland, Amsterdam (1986)
11. Gibson, J.J.: The Ecological Approach to Visual Perception. Houghton Mifflin, Boston (1979)
12. Johnson, M.: The body in the mind. In: The bodily basis of meaning, imagination, and reason, The University of Chicago Press, Chicago & London (1987)
13. Hacker, W.: Allgemeine Arbeitspsychologie [General Occupational Psychology]. Hans Huber, Bern (2005)
14. Hacker, W.: Action Regulation Theory and Occupational Psychology. Review of German empirical research since 1987. The German Journal of Psychology 18(2), 91–120 (1994)
15. Walliser, F.-S.: Entwicklung und Nachweisführung einer Methodik zur Einführung und Stabilisierung von veränderten Prozessen in der Produktentwicklung. [Development and Verification of a method for implementation and stabilization of changed processes in product development]. Dissertation (1999), http://archiv.tu-chemnitz.de/pub/1999/0020/
16. DIN EN ISO 9241-11. Ergonomic Requirements for office with visual display terminals – Guidance on usability. Beuth, Berlin (1998)
17. Zwahr, A. (ed.): Meyers Grosses Taschenlexikon [Meyers Big Pocket Encyclopedia], 10th edn. Bibliographic Institute, Mannheim (2006)

Creation of an Ergonomic Guideline for Supervisory Control Interface Design

Pere Ponsa and Marta Díaz

4all-L@b Usability Laboratory, GREC Research Group Knowledge Engineering,
UPC, Technical University of Catalonia
Av. Víctor Balaguer, s/n, 08800 Vilanova i la Geltrú, Spain
{pedro.ponsa, marta.diaz}@upc.edu

Abstract. In tasks of human supervision in industrial control room they are applied generic disciplines as the software engineering and the physical ergonomics for the design of the computing interface and the design of the control room layout. From the point of view of the human computer interaction, to these disciplines it is necessary to add the usability engineering and the cognitive ergonomics since they contribute rules for the user centered design. The main goal of this work is the creation of a cognitive ergonomic guideline for supervisory control interface design in order to improve the efficiency of the human machine systems in industrial automation.

Keywords: supervisory control, human-machine interface design.

1 Introduction

In recent years, control systems and the role of control room human operators have changed dramatically. Human operator activity has evolved from manually performing the process, to control system supervision. Today, the human operator requires an in-depth knowledge of the process that he/she is overseeing and the ability to make effective decisions within demanding constraints. The increased complexity of industrial process control calls for a new methodological approach (for research and design purposes), which reproduces the essential components of current control systems: the environment, the task at hand and human operator activity. The complexity of industrial process supervision makes it necessary to supplement the Human Factors approach and the Human-Computer Interaction approach with a cross-disciplinary cooperation in order to integrate knowledge and methods from other fields, especially Cognitive Ergonomics, Automation and Artificial Intelligence [1]. Our view is that complete control systems engineering must encompass all these approaches.

Ergonomics is concerned with the adaptation of technology to suit human operator need and ability so as to achieve effectiveness, efficiency and user/worker satisfaction and comfort.

Supervisory control is the set of activities and techniques developed over a set of controllers (programmable logic controllers and industrial regulators) which ensures the fulfilling of control goals. One of the main goals is to prevent possible plant

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 137–146, 2007.

malfunctions that can lead to economical lose and/or result in damage. For this reason, other fields of knowledge concerned with manufacturing systems performance – such as maintenance and industrial security – are complementary in the study of supervision systems.

In this paper a methodology for the creation of an ergonomic guideline for supervisory control interface design is proposed. In section 2 we present briefly the previous research on human interface design guidelines. A checklist of indicators of the guideline called 'ergonomic guideline for supervisory control interface design' (GEDIS Guia ergonómica para el diseño de interfaz de supervision in Spanish version) is described in section 3. In section 4, transition from the GEDIS model to industrial interface in control room is evaluated. Finally, conclusions and future research lines.

2 Previous Research

In this section we will briefly review the research on human interface design guidelines studies related to control and supervision tasks.

2.1 ISO 11064

The ISO 11064-7-2006 is a part of ISO 11064 that establishes ergonomic principles for the evaluation of control centers [2]. It gives requirements, recommendations and guidelines on evaluation of the different elements of the control center, i.e. control suite, control room, workstations, displays and controls, and work environment.

Exist a relationship with other standards like ISO 13407 human centered design processes and ISO 9242 ergonomic visual display design.

2.2 Human Factors Design Standards (HFDS)

The Human Factors Design Standard (HFDS) is an exhaustive compilation of human factors practices and principles integral to the procurement, design, development, and testing of Federal Aviation Administration (FAA) systems, facilities, and equipment of the United States [3]. For example, Chapter 9 is related about human-computer interface and a set of general principles of screen design.

2.3 Human Interface Design Review Guidelines (Nureg 0700)

The U.S. Nuclear Regulatory Commission (NRC) staffs reviews the human factors engineering (HFE) aspects of nuclear power plants in accordance with the Standard Review Plan (NUREG-0800). Detailed design review procedures are provided in the HFE Program Review Model (NUREG-0711).

As part of the review process, the interfaces between plant personnel and plant's systems and components are evaluated for conformance with HFE guidelines [4]. The document, Human-System Interface Design Review Guidelines (NUREG-0700, Revision 2), provides the guidelines necessary to perform this evaluation [4].

The review guidelines address the physical and functional characteristics of human-system interfaces (HSI). The NRC staff can use the NUREG-0700 guidelines to evaluate a specific design.

2.4 I-002 Safety and Automation Systems NORSOK

The I-002 SAS is an example of NORSOK standards from the Norwegian petroleum industry [5]. This standard covers functional and technical requirements and establishes a basis for engineering related to Safety and Automation System Design.

This standard shall be used together with I-001 "Field Instrumentation", I-005 "System Control Diagrams", Z-010 "Electrical, Instrumentation & telecommunication Installation" and S-001 "Technical Safety". The SAS Life Cycle Cost should be used as a criterion for the evaluation of the system. This includes engineering, commissioning, documentation, spare parts, and production loss during system repair and modifications/maintenance in the operational phase.

2.5 Man Systems Integration Standard (NASA-STD-3000)

This document provides specific user information to ensure proper integration of the man-system interface requirements with those of other aerospace disciplines [6]. These man-system interface requirements apply to launch, entry, on-orbit, and extraterrestrial space environments.

This document is intended for use by design engineers, systems engineers, maintainability engineers, operations analysts, human factors specialists, and others engaged in the definition and development of manned space programs. Concise design considerations, design requirements and design examples are provided.

3 GEDIS Guideline

The GEDIS guide is a method that seeks to cover all the aspects of the interface design [7]. From the initial point of view of strategies for effective human-computer interaction [8] applied to supervision tasks in industrial control room.

3.1 List of Indicators

The GEDIS guide consists of 10 indicators that seek to cover all the aspects of the interface design in the supervisory control domain [9]. The indicators are: architecture, distribution, navigation, color, text font, status of the devices, process values, graphs and tables, data-entry commands, and finally alarms (see Table 1 and Table 2). Navigation indicator of Fig. 1 shows a possible layout to locate all the connections between screens. Distribution indicator of Fig. 2 shows a possible layout to locate all the objects inside the screen. layout inside the screen for Distribution indicator.

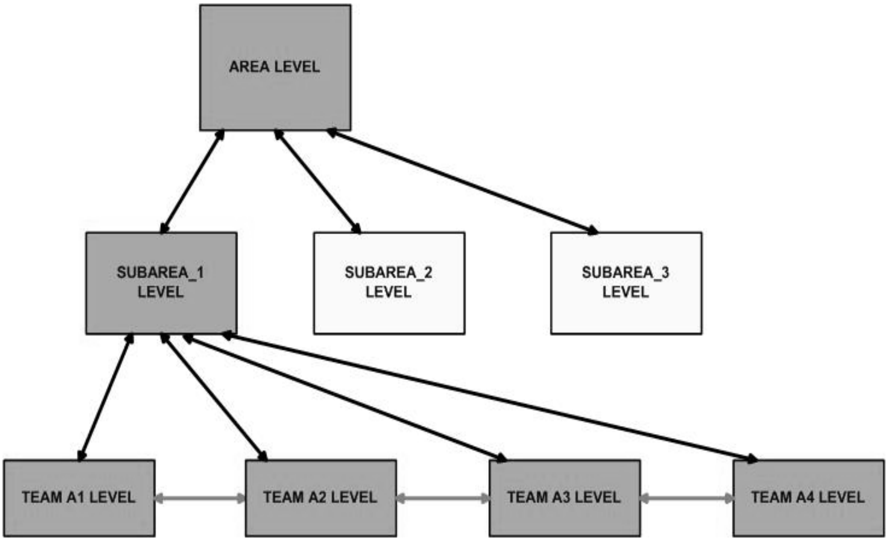


Fig. 1. A typical cyclic network menu in supervisory control interface associated to Navigation indicator

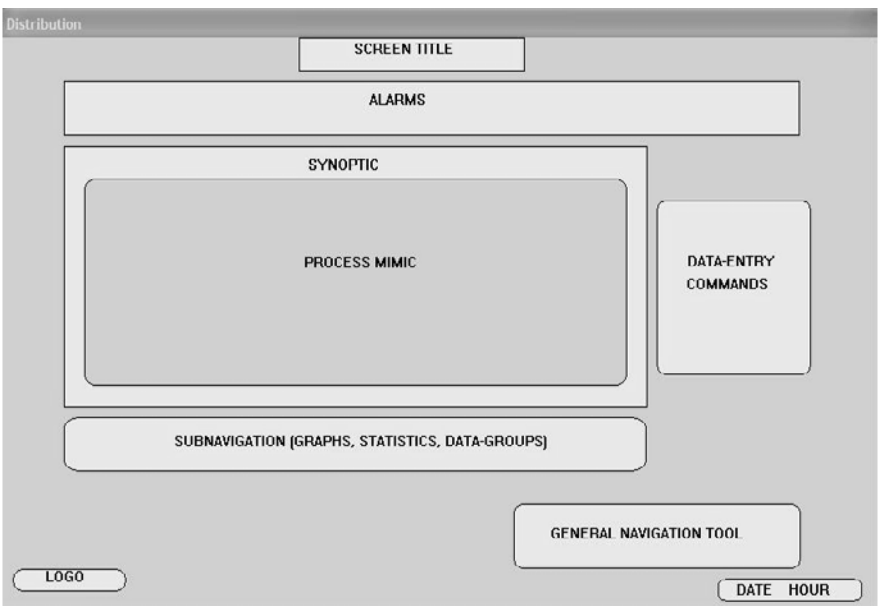


Fig. 2. An example of object's layout inside the screen for Distribution indicator

Table 1. GEDIS guide indicators (part one)

Name of indicator	Numeric qualitative range	
Architecture		
Map existence	[YES, NO]	[5,0]
Number of levels le	[le<4, le>4]	[5,0]
Division: plant, area, subarea, team	[app, med, no app]	[5, 3, 0]
Distribution		
Model comparison	[app, med, no app]	[5, 3, 0]
Flow process	[clear, med, no clear]	[5, 3, 0]
Density	[app, med, no app]	[5, 3, 0]
Navigation		
Relationship with architecture	[app, med, no app]	[5, 3, 0]
Navig. Between screens	[app, med, no app]	[5, 3, 0]
Color		
Absence of non appropiate combinations	[YES, NO]	[5,0]
Color number c	[4<c<7, c>7]	[5, 0]
Blink absence (no alarm situation)	[YES, NO]	[5,0]
Contrast screen versus graphical objects	[app, med, no app]	[5, 3, 0]
Relationship with text	[app, med, no app]	[5, 3, 0]
Text font	[f<4, f>4]	[5, 0]
Font number f		
Absence of small font (smaller 8)	[YES, NO]	[5,0]
Absence of non appropriate combinations	[YES, NO]	[5,0]
Abbreviation use	[app, med, no app]	[5, 3, 0]

where, app= appropriate, med= medium and no app= no appropiate.

3.2 Evaluation

The evaluation expressed in quantitative numeric form or in qualitative format it seeks to promote the user's reflection that stuffs the GEDIS guide by way of questionnaire, so that it picks up the use experience that doesn't end up being verbalized in many occasions.

Each one of the indicators of the Table 1 and Table 2 can substructure in diverse subindicators. For example, the indicator Color can be detailed in: absence of non appropriate combinations (5), number of colors (5), blink absence (no alarm situation) (5), contrast screen versus graphical objects (3), relationship with text (3).

For each subindicator it is recommended it is punctuated numerically in a scale from 1 to 5. In this example the number of subindicators of the indicator Color is J = 5 (see formula 1).

The formula necessary to calculate the numeric value of each indicator is the formula 1.

$$Indicator = \frac{\sum_{j=1}^{J} w_j Subind_j}{\sum_{j=1}^{J} w_j} \tag{1}$$

where, Subind= subindicator and w = weight.

The mean value that one obtains by the formula 1 with these numeric values is 4,2 . If it is rounded, the value is 4, so that to the indicator Color it is assigned the value 4 in this example, considering that each one of the subindicators has the same weight (w1 = w2… =wJ = 1).

Table 2. GEDIS guide indicators (part two)

Name of indicator	Numeric qualitative range	
Status of the devices		
Uniform icons and symbols	[app, med, no app]	[5, 3, 0]
Status team representativeness	[YES, NO]	[5,0]
Process values		
Visibility	[app, med, no app]	[5, 3, 0]
Location	[app, med, no app]	[5, 3, 0]
Graphs and tables		
Format	[app, med, no app]	[5, 3, 0]
Visibility	[app, med, no app]	[5, 3, 0]
Location	[app, med, no app]	[5, 3, 0]
Grouping		
Data-entry commands		
Visiblity	[app, med, no app]	[5, 3, 0]
Usability	[app, med, no app]	[5, 3, 0]
Feedback	[app, med, no app]	[5, 3, 0]
Alarms		
Visibility of alarm window	[app, med, no app]	[5, 3, 0]
Location	[app, med, no app]	[5, 3, 0]
Situation awareness	[YES, NO]	[5,0]
Alarms grouping	[app, med, no app]	[5, 3, 0]
Information to the operator	[app, med, no app]	[5, 3, 0]
Information to the operator	[app, med, no app]	[5, 3, 0]

where, app= appropriate, med= medium and no app= no appropiate.

Each one of the indicators of the Table 1 is measured in a scale from 1 to 5. The human expert operator prepares in this point of concrete information on the indicator, so that it can already value the necessities of improvement. The values of the indicators can group so that the GEDIS guide offers the global evaluation of the interface and it can be compared with others interfaces.

The formula necessary to calculate the GEDIS guide global evalutation index is the formula 2.

$$Global\_evaluation = \frac{\sum_{i=1}^{10} p_i ind_i}{\sum_{i=1}^{10} p_i} \tag{2}$$

where, ind = indicator and p = weight.

In a first approach it has been considered the mean value among indicators expressed in the formula 2. That is to say, to each indicator it is assigned an identical

weight (p1 = p2… =p10 = 1) although it will allow it in future studies to value the importance of some indicators above others. The global evaluation is expressed in a scale from 1 to 5. Assisting to the complexity of the systems of industrial supervision and the fact that an ineffective interface design can cause human error, the global evaluation of a supervision interface it should be located in an initial value of 3-4 and with the aid of GEDIS guide it is possible to propose measures of improvement to come closer at the 5.

4 Study Case

By way of experience pilot the GEDIS guide has been applied to a real environment of supervision: the supervisory control interface of the Sugar Technology Center. This case is an example of Spanish industrial application.

4.1 CTA Simulator

The Sugar Technology Center (CTA Centro Tecnología Azucarera in Spanish version) uses the simulation for the training of control room personnel.

Sugar production is a complex process. It includes several production sections and many production units are involved, both for continuous and bath operation [9]. Hundred of process variables must be monitored and controlled, so a Distributed Control System DCS must be used and a set of model predictive controllers must be used.

The tasks of the human operators in this control room are: detection of anomalies in the production process and process operation. In some cases the human operator does not know enough about the process he is supervising. A way to prevent these problems: the Sugar Technology Center use expert systems for failure detection and diagnostics (predictive control algorithms).

The Sugar Technology Center has been developed a training simulator to modeling and simulating the production process and the human operators' supervisory tasks. The simulator developed in the CTA is an example of full scale simulator, a type of simulator that reproduces the whole operating environment [10]. This simulator emulates the control room of a sugar mill. A series of object oriented modelling library tools are used to create each part of the sugar mill: diffusion (see Fig. 3), evaporation, purification, sugar room, boilers, dryer, and liquor storage.

In this practical case the designer has generated a group of screens of which it has facilitated a sample. The GEDIS guide has been applied to posteriori in an external way and without the designer's collaboration. In a generic way by means of the use of the GEDIS guide has been detected a group of anomalies, they have intended solutions, it has been quantified each one of the indicators numerically and the global evaluation of the guide has been obtained for the studied interface. All the information has been sent to the CTA partners so that they value the possibility of design of some parts of the interface with the suitable improvements.

Concerning the first three evaluated indicators, an architecture composed by 3 layers is observed clearly, so that in the supervision interface the navigation prevails in not very deep width, aspect that is typical in the context of industrial interfaces design. As for the navigation ways among screens have been carried out corrections

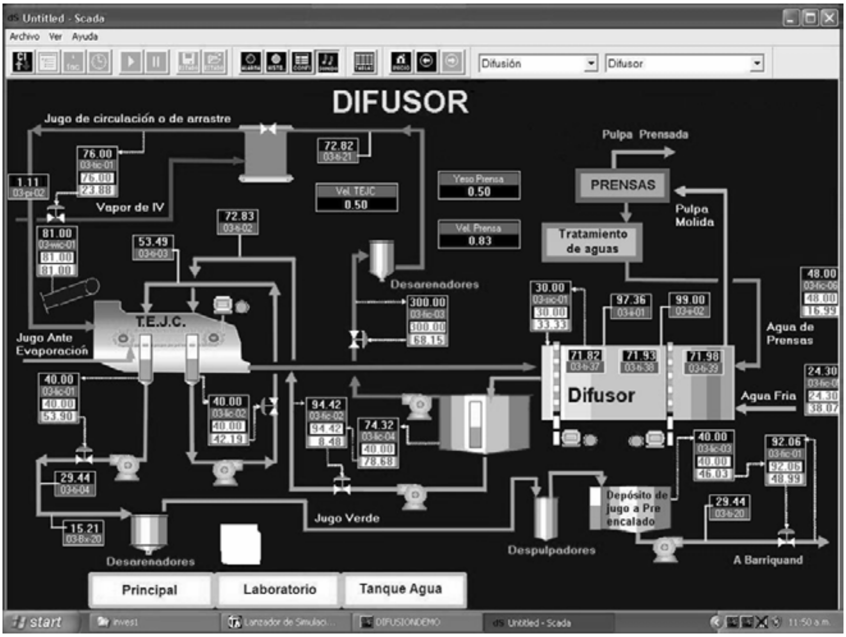


Fig. 3. An example of screen of the CTA simulator: diffusion team

since in the step from a screen to the following one the navigation submenus they change position and of format, disoriented to the user.

Regarding the use of the color the necessity has been commented of standing out by means of color the state of some components, as the distributed valves, so that it can be distinguished with clarity and differentiate the states open/close. In some screens a red and green excessive use of colors is detected in the outlying part. The GEDIS guide recommends the use of the red color associated to the alarm indicator.

Regarding the graphics of tendencies and tables, although a clear graphic representation has been observed of each one of the variables and the control action, its throws in lack a representation of variables contained in an only one historical that allows the operative to evaluate the future tendency of this variables and to make some decision (changing the set point, changing the controller's parameters).

The global evaluation that the GEDIS guide makes of the supervisory control interface of the CTA simulator it is located in 3,5. By means of the mentioned corrections, the index can arrive without problems between 4 and 5, that is to say the maximum values of the numeric scale.

5 Conclusions

In tasks of human supervision in industrial control room is habitual that an external designer, - by means of the commercial programs Supervisory Control and Data Acquisition SCADA -, take charge of designing the supervision interfaces in function to the knowledge on the physical plant and the group of physical-chemical processes

contributed by the process engineers. Although standards exist about security in the human machine systems that impact in aspects of physical ergonomics, interface design by means of rules of style, it is remarkable the absence of the design of interactive systems centered in the user where the engineering usability and the cognitive ergonomics can contribute significant improvements [11], [12], [13], [14].

The GEDIS guide is an approach that tries to fill a methodological hole that joins the efforts of the systems engineering and the ergonomics for the improvement of the effectiveness of the human-machine system in industrial control room.

The application of the GEDIS guide to the study of cases contributes among other details the measure in form of indicators of aspects of interface design, the recommendation of changes for the improvement of the interface, and a global evaluation index that allows quantifying the current state of the interface regarding the future state after applying the correct measures. The studied case presented shows a Spanish industrial application. With the GEDIS guide approach it's possible to perceive diverse anomalies and to propose improvements in the interface design.

In these moments our usability laboratory is analyzing the GEDIS guide to simplify the number of indicators of the guide, to improve the evaluation method, and to promote the use of the guide inside the cycle of life of the software engineering, in this case in the early phases of the supervisory control interface design.

Acknowledgements. The authors thank César de Prada and Felipe Acebes, members of the Systems Engineering and Automatic Department of the University of Valladolid, Spain, their collaboration and kindness for the use of the Sugar Technology Center CTA simulator. This work has been partly supported by the Spanish MEC project ADA-EXODUS (DPI2006-15630-C02-01).

References

1. Samad, T., Weyrauch, J.: Automation, control and complexity. John Wiley and Sons, Chichester (2000)
2. ISO International Organization for Standarization: Ergonomic design of control centres, parts I, II, III, IV (2004) En URL http://www.iso.org
3. Federal Aviation Administration: Human factors design guide for acquisition of commercial-off-the-shelf subsystems, non-developmental items, and developmental systems (DOT/FAA/CT-96/01). Atlantic City International Airport, DOT/FAA Technical Center (1996)
4. Holstom, C.: Human factors and control centre issues. What lessons we have learned. Institute for Energy Tecnology, OECD Halden Reactor Project (2000)
5. U.S. Nuclear Rgulatory Commission: NUREG-0700, Human-system interface design review guidelines. Office of Nuclear Regulatory Research, Washington DC 20555-0001 (2002) En URL: http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr0700/nureg700.pdf
6. Norsok Standard: I-002 Safety automation system. Norwegian Technology Centre Oscarsgt. 20, Postbox 7072 Majorstua N-0306 Oslo. (2006) En URL: http://www.olf.no y http://trends.risoe.dk/detail-organisation.php?id=52#corpus
7. NASA: Man system integration standards, NASA-STD-3000 (1995) En URL: http://msis.jsc.nasa.gov/

8. Gallareta, J.G., Ponsa, P.: La ergonomía cognitiva aplicada al diseño de interfaces computerizadas persona-máquina en salas de control industriales. Documento interno del Departamento ESAII de la Universidad Politécnica de Cataluña (2005)
9. Schneiderman, B.: Designing the user interface. Strategies for effective human-computer interaction, 3rd edn. Addison-Wesley, Reading (1997)
10. Centro de Tecnología Azucarera: Simulador de sala de control de fábricas azucareras (2006) En URL: http://www.cta.uva.es/CTANEW/espanol/proyectos/Simuladores/Frame_Simuladores.htm
11. Merino, A., Pelayo, S., Rueda, A., Alves, R., García, A., Acebes, F., De Prada, C., Gutiérrez, G., García, M.: Un simulador de alcance total para la formación de los operarios de sala de control de factorías azucareras. XXIV Jornadas de Automática, León (2003) En URL: http://www.cea-ifac.es/actividades/jornadas/XXIV/documentos/si/200.pdf
12. De Arquer, I., Nogareda, C.: NTP 544 Estimación de la carga mental de trabajo: el método TLX. Instituto Nacional de Seguridad e Higiene en el Trabajo (2006) En URL http://www.mtas.es/insht/ntp/ntp_544.htm
13. Nimmo, I.: Designing control rooms for humans. Control Magazine (2004)
14. Granollers, T., Lorés, J., Cañas, J.J.: Diseño de sistemas interactivos centrados en el usuario. Editorial UOC, Colección Informática, 43 (2005)
15. Cañas, J.J.: Personas y máquinas. Editorial Pirámide, Colección Psicología (2004)

Ergonomists and Usability Engineers Encounter Test Method Dilemmas with Virtual Work Environments

Ari Putkonen and Ursula Hyrkkänen

Turku University of Applied Sciences, Sepänkatu 3, 20700 Turku, Finland
ari.putkonen@turkuamk.fi

Abstract. Today's ergonomists and usability engineers need a broad understanding of the characteristics and demands of complex sociotechnical systems in order to develop virtual work systems and mobile communication tools for workers. Familiarity with appropriate ergonomics tests and evaluation methods is a prerequisite of this understanding. The literature review about ergonomics methods was performed. Applicable, potential and inapplicable ergonomics test methods for virtual work systems have been identified, based on the validity analysis and case example. The large number of available methods is confusing for ergonomists and therefore a hierarchical top-down approach is needed for method selection. The issues highlighted in this paper may be useful for ergonomists and usability practitioners who are participating design processes in complex virtual work environments.

Keywords: ergonomics, work system design, human-centred design, virtual work, test methods.

1 Introduction

The capacity of workers to percept and process information is burdened with the complexity and high demands of working life. Knowledge of the complexity factors of the overall work system is essential for an in depth understanding of human working capabilities and limitations [17]. It is also essential for proper test methods selection during design process. The complexity of work is usually considered as a factor related to the task. At one end the task is creative and demanding and at the other end it is simple and routine-like [2].

Working environments are changing from the traditional model. An increasing amount of work takes place in networked and virtual environments which are not tied to one place and time. The work is defined 'mobile', if the employee works more than ten hours per week outside of the primary workplace and uses information and communication technology (ICT) for communication [9], [29]. The use of ICT tools generates the virtual work environment. The planning of working conditions becomes challenging, because there is a lack of proper tools for analysing and testing mobile working conditions.

Society is becoming more and more dynamic and complexity of work is increasing, this has several implications that cause new challenges to the ergonomists and usability engineers. Rapid develop in technologies along with economic demands

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 147–156, 2007.

have led to a noticeable increase in the complexity of engineering systems. Rasmussen [22] emphasises that ergonomic contributions should be rather proactive than only responding to identified problems, but especially, they should be based on overall models of complex socio-technical systems [6]. In complex socio-technical system the increasing number of operation levels requires also that broader circumstances are considered during the product design process.

This paper presents the ergonomic test method dilemmas related to complex virtual work environments from the perspective of the design process of ICT devices intend for use in such an environment. The questions to be studied are; (i) How the complex virtual work environment can be described as a hierarchical work system, in order to support human-centred design approach and ergonomics method selection? (ii) What are the applicable ergonomic methods to use in complex virtual environment?

This article is organised as follows. The first chapter introduces the principles of the complex system design, the complexity factors of virtual work, the known dilemmas of method selection and the abstraction hierarchy model applied in this study. Next the methodological and empirical context of this study is presented. Thereafter, the article describes the results of the empirical cases illustrating the complexity of mobile work done in virtual spaces. Finally, the results of the article and some suggestions for ergonomic methods selection are summarised and discussed.

2 Development Methods of Complex Systems

In the 1960s Christopher Alexander and Herbert Simon developed early theories about how to design complex systems [7]. Since then, several design frameworks have been presented in theoretical and applied literature. The fundamental conclusion of these studies is that the larger work systems have to consider when there is a need to understand human-technology interaction, capabilities and limitations better [17].

2.1 Design Principles

Human-centred design is a design method for complex systems that addresses such problems by focusing mainly on the user [21]. There are some principles how the human-centred design approach can be realised. Norman [21] defines it as a process which starts with a multidisciplinary team that includes members from marketing, technology and user of product. The first task is to determine the product. According to Norman this seems to be obvious, but it is the most commonly ignored or badly examined task. Based on a thorough task analysis, the design process continues with the human-centred activities, like usability engineering [20]. The human-centred design process for interactive systems is formally described in ISO standard 13407 [13]. It formulates that human-centred design as a multidisciplinary activity, which incorporates human factors and ergonomics knowledge and techniques to enhance effectiveness and productivity, while improving human working conditions.

Task analysis is the process of analysing the way humans perform their jobs: the things they do, the things they act on and the things they need to know. This process will identify and document the user's tasks and significant user attributes. Overall

analysis of the user's tasks is the foundation of a human-centred design method. Different task analysis techniques exist, e.g. Hierarchical Task Analysis [5], [23].

It has been emphasised by Ulrich and Eppinger [28] that system-level issues are critical when developing complex systems including many integrating subsystems and components. The architecture for the overall system has to be addressed at the system-level design phase. Ulrich and Eppinger [28] also remind that the experience of the environment of the product or the context in which the users work or live is essential. Otherwise irrelevant product features may be developed and solutions for users' real needs may never be discovered.

Macroergonomics is a top-down approach to the design of work system, where the design characteristics of the overall work system are carried to the design of human interfaces. There are three criteria what are essential for an effective work system design: (i) joint design purpose of personnel subsystem and technological subsystem, which should be developed simultaneously and supported by employee participation thorough the entire design process; (ii) humanised task approach concerned with human functions and tasks in the work system, prior to the decision to allocate tasks to workers or devices; (iii) consideration of the organisation's sociotechnical characteristics, which should be evaluated and integrated into the design process of work system. When the selected development methodology fulfils the above mentioned three criteria design is human-centred and macroergonomic [10].

2.2 Virtual Work Environment and Its Complexity Factors

Working across organisational, geographical, cultural and temporal boundaries, increases the complexity of work systems [6], [12]. These demands are met especially in virtual work environments.

The concept 'virtual' is widely used in various frameworks. There is much discussion of virtual space, virtual groups and virtual organisation. Virtual space is used for communication and collaboration: it refers to electronic working environment where documents, messages and images and even avatars are stored, exchanged, retrieved and worked. Virtual group signifies a number of persons, who are to a certain extent dispersed in space and sometimes also in time, communicating through the media [3].

The complexity of work is usually considered as a factor related to the task [2]. The expanded complexity concept considers also the working environment that can be a different combination of physical, virtual, social and cultural spaces.

Vartiainen [29] has described the complexity of virtual team working contexts by applying the model of six complexity characteristics. The characteristics are mobility, geographical dispersion of the workplaces, diversity of actors, asynchronous working time, temporary structure of the working groups and mediated interaction. These six dimensions form in addition to task complexity a set of requirements that can also considered as ergonomic challenges for constructing virtual working space [19].

According to the previous studies the challenges of virtual work entails many novel questions for ergonomics discipline. For example working in virtual, geographically distributed manner has influence on working procedures, coordination and communication [8], [18]. It has also been proven that virtual, multi-seated work increases the physical distance of the workers of the main team and hinders face to face communication of the team. Non-verbal cues are absent in mediated communication

and this may easily lead to misunderstandings and lack of trust. Especially the global groups have members with different backgrounds, which may further create communication problems. The temporary nature of projects leads to loose social engagement due to the limited expectations of working together again. Asynchronous work time in relation to the main team makes further demands on communication. On account of this, the proper functioning of ICT technologies is compulsory.

2.3 Dilemma of Ergonomics Methods Selection

Ergonomics methods have been classified by various aspects by authors. They are, for instance, either evaluative or analytic [4], empirical or non-empirical [14], expert or non-expert performed, time consuming or fast, expensive or inexpensive and they can relate to different stages of the design process [24]. The one aim of classification has been to support the selection of applicable methods in design process.

Table 1. Summary of the ergonomics methods selected for consideration

Method	Overview/Objective	Type of approach	Work system context to consider
Checklists	Design evaluation and finding of improvements	evaluative	Human-Device interactions
Heuristics	Design evaluation and finding of improvements	evaluative	Human-Device interactions
Layout analysis	Examining of display and control layouts, device optimisation	evaluative	Human-Task interactions
Questionnaires	Predicts user satisfaction and perception	evaluative	Human-Task interactions
Hierarchical task analysis, HTA	Describes the task in terms of a hierarchy of operations	analytic	Human-Task interactions
Focus groups	Participates users and customers in discussion	evaluative	Work Process interactions
Observations	Expert observes users as they work in real work context	evaluative	Human-Task interactions
Error prediction methods	Systematic human error reduction and prediction	analytic	Human-Task interactions
Repertory grids	Predicts user satisfaction and perception	evaluative	Human-Device interactions
Link analysis	Examination of the way humans use displays, device optimisation	evaluative	Human-Task interactions
Keystroke level model, KLM	Measures speed of performance	analytic	Human-Task interactions
Interviews	Obtains in-depth data about a particular process or tasks	evaluative	Work Process interactions
Walkthrough	User do and explain demonstration of a task in realistic environment	evaluative	Work Process interactions
Macroergonomic analysis and design, MEAD	Framework for conducting work system improvements	analytic	Organisational interactions
Participatory ergonomics	Employee involvement in their own work activity design	evaluative	Organisational interactions

Some selection guidelines have been proposed in the literature as well. For instance, Kjeldskov and Skov [16] state that more realistic test environment and more experienced subjects discover more problems from the system. Stanton [26] states that the methods have to be selected by the required output of the test or analysis, choices are; errors, performance, usability or design characteristics. Hendrick and Kleiner [10] emphasise that early observation of the system's complexity is an increasingly important managerial task in order to design a work system where the well-being of workers and the overall system performance are in balance.

The fifteen ergonomics methods selected for consideration in this study are listed in Table 1. For each method the following information is given: an overview with an indication of objective, type of approach and typical context to use in the work system. The methods were selected based upon the patterns of usage [27], methods for macroergonomics applications [11] and our analysis, so that these are a representative variety of ergonomics methods to cover the analyses and evaluation needs of complex work system. Besides of above mentioned sources the details of these methods are presented in a number of ergonomics text books [15], [26].

2.4 The Abstraction Hierarchy as a Framework

The Abstraction Hierarchy (AH) is one of the best known representation frameworks describing complex work environments and adaptive sociotechnical systems [22]. The AH describes a system at different levels of abstraction using how and why relationships. Moving down the model levels answers how certain elements in the system are achieved, whereas moving up reveals why certain elements exist. Elements at highest level of the model define the purposes, goals and constrains of the system. Elements at the lowest levels of the model indicate and describe the physical forms (e.g. ICT device) of the system.

3 Method and Data Collection

The aim of this article is to depict how the well-known methods of ergonomic assessment function in virtual work sets. The question can be placed in the field of validity research. Validity refers to the degree to which a study accurately reflects or assesses the specific concept that the researcher is attempting to measure. While reliability is concerned with the accuracy of the actual instrument of measurement or procedure, validity is concerned with the study's success at measuring what the researchers set out to measure. Especially this study was directed to first steps of building an assessment framework for exploring construct and content validity of the ergonomic evaluation methods [1], [25].

First the assessment tool was constructed by defining the hierarchical model for complex, virtual work. This was done by using the AH as a theoretical framework for creating a model of the virtual work structure. Articles concerning the definition of virtual work environment, ergonomics methods and complex systems were extracted from databases. After that, expert proceeding was used to analyse articles and to construct a frame of reference applicable for describing the hierarchical system of virtual work and it's environment.

A single case study was used for testing the developed framework. The case was selected in cooperation with a company, operating some 800 service centres in more than

40 countries, the company's call centres are linked with the service offices and 13,000 employees. The selected group was one of their mobile maintenance groups (N=8). Information of this case was gathered by observing the work of theirs and by individual interviews. The data was coded with the assistance of the developed framework and classification. After the framework and classification tool was acceptable the existing and commonly known ergonomic evaluation methods were set against it.

4 Results of the Study

The studied virtual work organisation is described as a hierarchical work structure in Table 2. The abstraction hierarchy levels from the physical form to the functional

Table 2. Abstraction hierarchy of the virtual work system of the mobile workers

Whole/Part Means/Ends	Individual, mobile worker	Functional unit, service team	Sub-system, service centre	Total system, corporation
Functional purpose	Keeps up the machine	Accomplish customer service and maintenance on particular geographical region	Customer service and maintenance of products world wide	Worldwide business operations with industrial products
Abstract function	Service and maintenance of specific machines	Service and maintenance of the products	Plan and sustain availability, performance, efficiency and safety of the products	Manufacturing, marketing and service
Generalised functions	Planning of the service operations and travel logistics	Allocation of work force, job scheduling, improving of efficiency and quality	Worldwide service execution, customer management, control of safety requirements and legislation	Conducting strategy, running business, planning operations, coordination of units
Physical function, process and activity	Performs service tasks, utilises information and communication tools, travels	Data transfer between domains, decision making, customer contact management	Proactive service planning, negotiations and agreements with customers, invoicing	Customer relation management, financing planning, target setting, capacity planning
Physical form	Service in domain context, machines, tools, computers, documents, reports	Joint meetings in various places, shared calendar and tasks, work orders, reports	Office facilities, technical documentation, data network, servers, databases	Office facilities, data network, mainframes, databases

purpose interrelate with different sub-systems of organisation (whole/part), i.e. individual, functional unit, sub-system and total system. The smallest work system is individual as a mobile worker in this case. The functional unit represents a service team, which consist of some individuals. The service centre of the corporation is such a sub-system. Corporation level stands for the total system of the case organisation.

Vertically moving on the work system levels (means/ends) shows the how and why relationships between levels. Upper element answers why lower element exists and lower element tells how upper element is achieved. At the bottom mobile workers service machines in the domain context. At the next level, managers are planning operations and supplying resources to the mobile workers. Variety of managerial plans and functions are needed to maintain operations of sub-systems and, finally, the purpose of corporation is setting priorities through strategies and company policy.

<table>
<tr><td>Whole/Part
Means/Ends</td><td>Individual, mobile worker</td><td>Functional unit, service team</td><td>Sub-system, service centre</td><td>Total system, corporation</td></tr>
<tr><td>Functional purpose</td><td rowspan="5" colspan="4">Organisational interactions
MEAD
Participatory ergonomics

Work Process interactions
Interviews
Focus groups
Walkthrough

Human-Task interactions
Questionnaires
Hierarchical task analysis
Error prediction methods
Observations
Link analysis
Layout analysis
Keystroke level model

Human-Device interactions
Checklists
Heuristics
Repertory grids
Layout analysis</td></tr>
<tr><td>Abstract function</td></tr>
<tr><td>Generalised functions</td></tr>
<tr><td>Physical function, process and activity</td></tr>
<tr><td>Physical form</td></tr>
</table>

Fig. 1. Applicability of known ergonomics methods related to the work system's abstraction and organisational levels

Figure 1 depicts the abstraction hierarchy of the virtual work system as described above completed with the classified ergonomics methods. The upper the hierarchy level and the larger the system the more complex is the environment. The methods are

classified to four categories: human-device, human-task, work process and organisational interaction. The methods are not in certain order inside the category. The figure emphasises the hierarchy level between categories as indicating applicability to cover ergonomic development challenges in virtual work environment. Methods related the human-device and human-task interactions do not cover the development needs of the total system nor the ergonomic situation of the corporation. The methods in the work process and organisational categories are more potential and appropriate in virtual work environment.

5 Discussion

This paper aimed to answer for two questions about ergonomics methods and their use in complex virtual work environment. First the complex virtual work environment was analysed as a work system with the reference the case organisation. A frame of reference applicable to describing the hierarchical work system in virtual work was constructed with help of the AH –model. This new model supports the human-centred design approach and ergonomics method selection by the comprehensive approach.

The applicable ergonomic methods for complex virtual work environment assessment were identified by classifying them according to the extent of work complexity. They were classified to categories of human-device, human-task, work process and organisational interaction. The methods in work process and organisational categories are more applicable in virtual work environment than the methods considering human-device or human-task interactions. Methods related the human-device and human-task sub-systems only do not clarify the ergonomic situation of the total system nor the ergonomic situation of the corporation. Anyhow, human-device and human-task type of methods may also be valid in virtual work environment. It requires that the ergonomic analysis is first performed at upper functional levels and the obtained consequences are further derived to the specific sub-system as source data.

As organisations emerge towards virtual work environments, the need for advanced design methods will increase. However, analysing, measuring and developing the ergonomics of virtual environments is a difficult assignment. Identifying complexity factors of these environments and recognising limitations of existing design methods is a good start. The existing selection recommendations of ergonomics methods are mainly based on the concept of traditional organisational work. This study contributes to the discussion by offering one approach for selecting the ergonomics methods for virtual work environment.

When the functioning of the entire virtual working environment is under development, a human-system interaction and the entire work organization and socio-technical system have to be taken jointly into consideration. Macroergonomics is an approach of work systems design which attempts to achieve a fully harmonised work system. In the future, macroergonomics should be more common knowledge among ergonomists and usability engineers in order to meet the design and development challenges of complex work environments. This means, for example, that workers should be more involved in the design and implementation of technology and new information and communication systems in organisations. Our paper can hopefully highlight this significantly important issue in the future and encourage researchers for further studies.

References

1. Adcock, R., Collier, D.: Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. American Political Science Review 95, 529–546 (2001)
2. Andriessen, J.H.E.: Working with groupware. In: Understanding and evaluating collaboration technology, Springer, Heidelberg (2003)
3. Andriessen, J.H.E., Vartiainen, M.: Emerging mobile virtual work. In: Andriessen, J.H.E., Vartiainen, M. (eds.) Mobile Virtual Work. A New Paradigm?, Springer, Heidelberg (2006)
4. Annett, J.: A Note on the Validity and Reliability of Ergonomics Methods. Theoretical Issues on Ergonomics Science 3, 228–232 (2002)
5. Annett, J., Stanton, N. (eds.): Task analysis. CRC Press, UK (2000)
6. Carayon, P.: Human Factors of Complex Sociotechnical Systems. Appl. Ergon. 37, 525–535 (2006)
7. Clark, K.: Design rules: The power of modularity. MIT Press, Cambridge (1999)
8. Duarte, D.I., Snyder, N.T.: Mastering Virtual Teams. Strategies, Tools and Techniques that Succeed (2001)
9. Gareis, K., Lilischkis, S., Mentrup, A.: Mapping the mobile eWorkforce in europe. In: Andriessen, J.H.E., Vartiainen, M. (eds.) Mobile Virtual Work. A New Paradigm?, pp. 45–69. Springer-Verlag, Berlin (2006)
10. Hendrick, H.W., Kleiner, B.M.: Macroergonomics: An introduction to work system design. Human Factors and Ergonomics Society, Santa Monica (2001)
11. Hendrick, H.W. (ed.): Macroergonomics: Theory, methods and applications. Lawrence Erlbaum Associates, Mahwah, NJ, USA (2002)
12. Hyrkkänen, U.: Analysis of work load factors and well-being in mobile work. In: Vartiainen, M. (ed.) Workspace Methodologies - Studying Communication, Collaboration and Workscapes. Report, /3. Helsinki University of Technology, Laboratory of Work Psychology and Leadership, Espoo, pp. 63–79 (2006)
13. International Organization for Standardization: Human-Centred Design Processes for Interactive Systems (1999)
14. Jordan, P.W.: Usability and product design. In: Karwowski, W. (ed.) International Encyclopedia of Ergonomics and Human Factors, pp. 1426–1428. Taylor & Francis, Florence, KY, USA (2001)
15. Karwowski, W. (ed.): InInternational Encyclopedia of Ergonomics and Human Factors. Taylor & Francis, Florence, KY, USA (2001)
16. Kjeldskov, J., Stage, J.: New Techniques for Usability Evaluation of Mobile Systems. Interact Comput 60, 599–620 (2004)
17. Kleiner, B.M.: Macroergonomics: Analysis and Design of Work Systems. Applied Ergonomics 37, 81–89 (2006)
18. Kokko, N., Vartiainen, M., Hakonen, M.: Work Stressors on Virtual Organizations (2004)
19. Nenonen, S., Gersberg, N.: Multidisciplinary workplace research. In: Vartiainen, M. (ed.) Workspace Methodologies - Studying Communication, Collaboration and Workspaces. Helsinki University of Technology, Laboratory of Work Psychology and Leadership, Report 3, Espoo, p. 93 (2006)
20. Nielsen, J.: Usability engineering. Morgan Kaufman, San Francisco (1994)
21. Norman, D.: The invisible computer. MIT Press, Cambridge (1998)
22. Rasmussen, J.: Human Factors in a Dynamic Information Society: Where are we Heading? Ergonomics 43, 869 (2000)
23. Shepherd, A.: Hierarchical task analysis. Taylor & Francis, New York (2001)

24. Stanton, N., Young, M.: A survey of ergonomics methods. In: Karwowski, W. (ed.) International Encyclopedia of Ergonomics and Human Factors, pp. 1903–1907. Taylor & Francis, Florence, KY, USA (2001)
25. Stanton, N., Young, M.: Utility analysis. In: Karwowski, W. (ed.) International Encyclopedia of Ergonomics and Human Factors, p. 1163. Taylor & Francis, Florence (2001)
26. Stanton, N., Young, M.: A Guide to Methodology in Ergonomics: Designing for Human use (1999)
27. Stanton, N., Young, M.: Is Utility in the Mind of the Beholder? A Study of Ergonomics Methods. Applied Ergonomics 29, 41–54 (1998)
28. Ulrich, K.T., Eppinger, S.D.: Product design and development, 3rd edn. McGraw-Hill Companies, Inc, New York (2004)
29. Vartiainen, M.: Mobile virtual work - concepts, outcomes and challenges. In: Andriessen, J.H.E., Vartiainen, M. (eds.) Mobile Virtual Work. A New Paradigm?, pp. 13–44. Springer-Verlag, Berlin (2006)

Interactive Style of 3D Display of Buildings on Touch Screen

Weina Qu and Xianghong Sun

State Key Laboratory of Brain and Cognitive Science, Institute of Psychology,
Chinese Academy of Sciences,
Beijing 100101, China
{quwn,sunxh}@psych.ac.cn

Abstract. This study was concentrated on the effect of different building 's floor number display modes and multi-speeds of view changing, for a firefighter recognized the fire scene through the 3D interface display system using a touch screen .The result showed that : (1) Showing number on floors is the worst way. showing numbers on wall, and on two sides of building whatever fixed, or rotating with the building were better than the way of cube display.(2) It's found the condition of "automatic rotation at fixed speed" at 30s was the most helpful for the decision efficiency, "automatic rotation with optional choice" the least helpful. Some mechanism and implications were discussed.

Keywords: Fire alarm system; 3D interface display system; Floor-numbering design; 3D display; Rotation.

1 Introduction

Fire alarm system is an essential part of high buildings in modern times, which helps firefighters' detection more efficiently and reduces the casualty [1]. The challenge, obviously, is not only make information available to people at any time, at any place, and in any format, but provide "proper" information, at "right" time, in "right" format [2,3]. In order to help the firemen to detect the original fire and the current fire spread in a high rise building as quickly and accurately as possible, 3D interface display system, as an important cognitive model [4,5] ,is necessary for it's convenience [6]. For the convenience of firefighters' decision-making and operation, 3D interface is easier to get the whole spread of fire, especially for vertical spread, very useful for recognizing the features of fire, and diagnosing the cause of fire, also useful in finding the way to approach the fire by the experiments we conducted. But it still had many things need to be improved in some aspects. For example, 1) not very useful to find the starting point of fire because there was not any information to show the location of the fire in the building; 2) Can't tell which floor the fire is located. Have to take long time to count 3) Can't see the details of each floor.

In order to improve 3D display, floor numbering, and more controls were added in a new prototype. How to show floor numbers and how to show the whole building, the building shape and details by 3D display were the main issues in this study. For

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 157–163, 2007.

the 3D building display, the horizontal rotation ways and rotation speeds, and the ways of building tilting were tested and discussed.

2 Method

There are many different cognitive model in the interact between man and environment, different job-domain's characters and different control mode's requirements create the different interaction cognitive model [7]. To got the best mode of the 3D interface to fit the firefighter, we did these two experiment do detect many 3D display modes.

2.1 Experimental Environment

A new FID prototype (Version 2.0) has been developed as the experimental environment. All the fire scenarios were supposed to happen in 3 buildings.

Floor numbers were shown in 3D display in 5 different optional ways, and a new control for building rotation was added. Right now there are three ways to show building shape and detail by rotating, i.e. 1) rotate continuously at 2 optional speeds, 2) rotate step by step, and 3) rotate continuously and automatically, which means there was no way to control the rotation speed.

2.2 Experimental Participants

9 Fire firefighters aged from 24 to 28 years participated in our experiment. One of them has bachelor degree and one-year firefighting experience, and others have more than 5-year experience.

2.3 Experiment 1

2.3.1 Experimental Design

Five floor-numbering designs were tested in our test, the first way is putting floor numbers on the building wall; the second is putting numbers on two sides of the high rise; the third is putting them on building floors; the fourth is putting them on two sides of the high rise but the location of floor numbers keeps still while the building display keeps rotating; the fifth is similar as the second, but the floor numbers were shown in 3D mode. The purpose of this study was to compare the five designs and find which way is more efficiency and more favorable to firefighters.

2.3.2 Experimental Tasks

The hospital building and the tower building were used as stimulus background. Participants were asked to tell the floor number.

2.3.3 Procedure

The hospital building and the tower building were used as stimulus background. Each type of floor numbering was shown on one building. There were two test sequences of floor numbering shown in the table 1.For each experimental condition, several highlighted floor levels would be shown on the PC screen one by one. Participants were asked to tell the fired floor number as accurate and quickly as possible.

Table 1. Two test sequences of floor numbering

Stimulus order 1	Stimulus order 2
Wall—Hospital building	Fixed—Hospital building
Side—Tower building	Side—Tower building
Floor—Hospital building	Floor—Hospital building
Fixed—Tower building	Wall—Tower building
Cube—Hospital building	Cube—Hospital building

From the prototype interface, there were several ways to know which floor was currently highlighted. In order to avoid participants using other cues to find it, the left part of PC screen was masked by a piece of paper. And participants were asked to answer the floor number only by watching the rotating building.

2.3.4 Experiment 1 Results

Table 2 showed the percentage of correct responses and response time. It's found that showing number on floors was the worst way. It got the most errors, and took the longest time. (For the reaction time, F(4)=6.899, p=0.00; for the percentage of correct answer, F(4)=7.939, p=0.00.) Among the other four ways, there was no significant difference existed. But showing numbers on wall, and on two sides of building whatever fixed, or rotating with the building were better than the way of cube display.

Table 2. Percentage of correct responses and response time

	Wall	Side	Floor	Fixed	Cube
Percentage of correct answers	95%	93%	52.8%	97.8%	89.6%
Reaction time (ms)	1412.5	1555.5	11311.1	1344.4	2077.8

The fig 1 showed firefighters' preference to the five kinds of floor numbering designs. It implied that firefighters' preference was consistent with the results of reaction time, and the percentage of correct answer. Showing numbers on floor got the least favorites. Cube way got the second least, and the other three were similar.

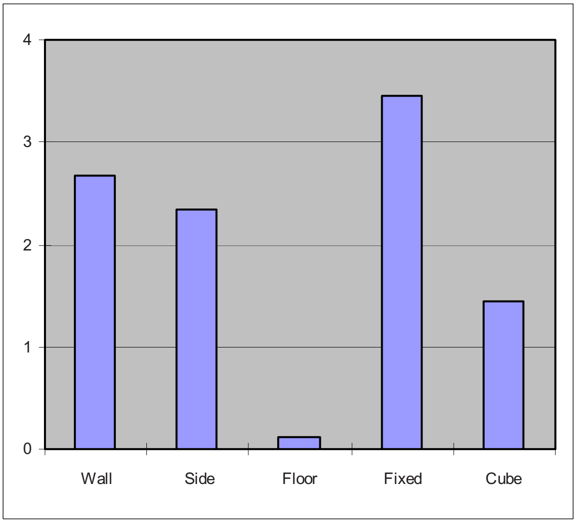


Fig. 1. Subjective ratings to different ways of floor numbering

From the three indexes we have got above, we would conclude that showing numbers on wall, or on two sides of building were better ways than showing them on floors. Going back to the behavior record, we found number shown on the floor was difficult to recognized. That's the reason people took longer time to tell the floor number. The reason why the first, second, and fourth ways got the similar performance could be that, because of the unregulated shape of hospital building, current interface was not able to show number on wall when the floor level was higher than 3. That made the three ways of showing number almost the same, so the performance and the subjective ratings under these three conditions were similar.

2.4 Experiment 2

2.4.1 Experimental Design

Right now the most popular way to show whole building in 3D mode is rotation. With the building rotating, people can see the building from different point of view. Nothing can be covered. There are two different rotation styles: one is rotating smoothly and continuously; the other is rotating step by step at a constant angle. People like building rotating continuously, but when the building structure is complex the programming costs of 3D rotation becomes much higher. So rotating step by step for high rise building is an alternative way to show all the structure details of it.

In this experiment, the rotation way was the independent variable including three levels: rotating continuously at two optional speeds, rotating step by step at two optional rotation angles, and rotating continuously at a fixed speed. First way allowed

people choose one rotation speed from 1to10, which was called rotation “automatically”; Second way allowed people choose one rotation angle: 15o, or 30o,and rotate the building by clicking the 15o, or 30o button, which was called rotation “manually”; Third way was similar as the first one, but people were not allowed to control the rotation speed. It was fixed at 5. The third was called “none” in the setting menu of current prototype, which meant “automatically rotation at fixed speed”.

2.4.2 Experimental Tasks

Each participant were asked to put out 18 fires including 6 single fires spread on single floor, 6 single fires spread on several floors, and 6 multi-fires (that meant there were several fire seeds in the building simultaneously). Half of the 18 fires were shown in 3D mode, the others shown in 3D with 2D floor plan.

For the way of rotation, 6 of the 18 fire scenarios were shown randomly in “auto” mode; 6 of them were shown in “manually” mode; the other 6 were shown in “automatic rotate at fixed speed” mode.

2.4.3 Procedure

For each fire scenario, participant was asked to accomplish three tasks: 1) find the starting point of the fire; 2) find the spread way; 3) find the path in the building for firefighter to put out the fire. Performance was recorded at three time points: 30 seconds, 90 seconds, and 5 minutes from the beginning when the fire was explored to participant. At each time point, the screen was suddenly black, and all the information was masked. Participant was asked to answer three questions: where was the fire? How did it spread? And how did you get to the fire to put it out? We counted the number of how many tasks were finished. Each of the three tasks was treated as one score. For example, if the participant only found the starting point of the fire at 30s time point, then found the spread of fire and the path to the fire at 90s time point, the participant would get 1 score at 30s, 3 scores at 90s, and also 3 scores at 5mins.

The reason why we recorded the performance at three time points was that, it’s quite urgent when firefighters made a decision during a fire. Decision was usually made within one minute. If the interface prototype was good enough, it must be helpful at the first one, or two minutes. Therefore we set three time points: the first two time-points were for the measurement of efficiency, the third one was trying to investigate that, if firefighters had enough time, what kind of information they looked for to adjust, or confirm their decision.

2.4.4 Experiment 2 Results

2.4.4.1 The Way of Rotation. The way of rotation Table 3 showed how many scenar-ios, under three kinds of rotations, were finished at each time points. In the volume of task performance, ‘0’ meant none of the three tasks was finished; ‘3’ meant all of the three tasks were finished.

Table 3. Number of people who finished the task successfully under different rotating conditions

Time point	Task performance	Rotate			Total
		Auto	Manually	Auto (with fixed speed)	
	0	5	6	3	14
30s	1	22	12	12	46
	2	10	8	4	22
	3	17	23	30	70
Total		54	49	49	152
	0	2	1		3
90s	1	5	4	4	13
	2	9	3	1	13
	3	38	41	44	123
Total		54	49	49	152
	0	1			1
5min	1	3	1	1	5
	2		1		1
	3	50	47	48	145
Total		54	49	49	152

Comparing with the people numbers at three time points, it's found that, at 30s, participants didn't finish all the tasks; but at 90s, and 5mins, almost all the people finished. It proved that firefighters only took one, or two minutes to make decision.

Comparing the three rotation conditions at three time points, people finished all three tasks under situation of automatic rotation at fixed speed were more than people under the other two conditions at 30s, and 90s. From the chi-square test, it's found the condition of "automatic rotation at fixed speed" at 30s was the most helpful for the decision efficiency, "automatic rotation with optional choice" the least helpful (30s:X2(6) =11.057 , p= .087). There was no difference among the three conditions at 90s, and 5mins (90s:X2(6) =10.026 , p= .124 5min:X2(6) =5.304 , p= .505).

In this experiment, both the first and third conditions were automatic rotation. But the third one was better than "rotating manually", and the first worse. It looks strange. Only difference between the first and third condition was the 3D control. First condition allowed firefighters control the rotation speed, but third one didn't. It probably meant that the manual control part slowed down information acquirement and the process of decision making.

2.4.4.2 Subjective Ratings to the Three Kinds of Rotations. From Table 4 we found 4 out of nine firefighters preferred automatic rotation with speed choices, another 4 of nine preferred rotating manually, only one of them like automatic rotation at a fixed speed. For the rotation speed and rotation angle, the choices were diverse. So, people's preference is not consistent with their performance.

Table 4. Preference to three kinds of rotating

Horizontal rotation	Rotate automatically (speed is controlled manually)			Rotate manually			Rotate automatically (speed is fixed)
Number of people	4			4			1
Speed of rotation	1 – 5	5 – 10	10	15o	30 o	Both	
Number of people	1	2	1	2	1	1	1

3 Conclusion and Discussion

In two experiments, we compare the five designs on floor number and compare building rotation way about the 3D building. We can make the following conclusions:

(1) Showing number on floors is the worst way. It got the most errors and took the longest time. Among the other four ways, there was no significant difference existed. But showing numbers on wall, and on two sides of building whatever fixed, or rotating with the building were better than the way of cube display.

(2) It's found the condition of "automatic rotation at fixed speed" at 30s was the most helpful for the decision efficiency, "automatic rotation with optional choice" the least helpful. There was no difference among the three conditions at 90s, and 5mins.

References

1. Fang, Z.: Development of human-machine interaction: multimedia and multisensory. Human Factors 2, 34–38 (1998)
2. Li, X.: Research on Adaptive Human Computer Interface Doctoral Dissertation, Southwest Normal University, 1 (2004)
3. Everhart, A.: A Lesson in Campus Fire-Alarm System Design. Consulting-Specifying Engineer 1, 53–54 (2006)
4. Chen, Y., Bai, S., Liang, B.: HLA-based 3D Virtual Environment For Synthetic Simulation System. Computer Engineering and Applications 38, 21–24 (2002)
5. Wang, S.: Frame of Pattern Recognition Model Based on Cognitive Psychology. Editoral Board of Geomatics and Information Science of Wuhan University 27, 543–547 (2002)
6. Song, T., Liu, W.: Research of Cognitive-Based 3D Model for Visualization of Association Rule. Microcomputer Applications 1, 1–3 (2006)
7. Rasmussen, J.: Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE Transactions on Systems, Man and Cybernetics, SMC-13, 257–266 (1983)

The Role of Human Factors in Design and Implementation of Electronic Public Information Systems

Karl W. Sandberg and Yan Pan

Mid Sweden University, Department of Information Technology and Media, Holmgatan 10, SE 951 70 Sundsvall, Sweden

Abstract. Design of Electronic Public Information Systems (e-PIS) can benefit from the inclusion of guidelines that consider the needs of users. Inclusion of guidelines and prototyping practices that take into consideration the capabilities of human users make it possible to avoid errors that would otherwise significantly reduce efficiencies offered by Information Technology (IT). Design strategies that include consideration of the Human Factor (HF) within e-PIS are introduced and discussed. Argued is that without integration of design strategies that consider human capabilities the efficiencies brought about through the use of IT are significantly reduced. Through the study of intended user behaviours during the development of systems loss of efficiency is avoided.

Keywords: Human Factors (HF), Electronic Public Information Systems (e-PIS), Design and Implementation.

1 Introduction

Human Factors (HF) can be traced to early efforts by industrial engineers, psychologists, and efficiency experts to streamline manufacturing operations and equipment for better worker efficiency. After the World War II, Human factors analyses and tests became routine in the design of military and commercial, a special focus was cockpit design of aircraft, since bad design of controls and displays often induced pilot errors. There was a growing recognition that understanding the human aspect is as critical as IT-based solutions. The concept of Human Computer Interaction (HCI) came into being as a result of this recognition.

The early emphasis of HCI on physical aspects such as the improvement of console button, switch, and display designs predated regulatory requirements for user comfort and was concerned more with improving efficiency than user comfort/health. As the design of controls became more sophisticated, HCI specialists begin to recognize that user cognition is an important design part. Design studies such as how controls were arranged and information was presented assessed both physical ability and average user cognition. From this trend, the specialities of cognitive engineering and usability emerged. The two are generally understood to mean the study of human interaction with complex interactive devices and systems with the ultimate goal being to ensure that people can carry out interactive tasks easily, effectively, and in a satisfying manner [1].

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 164–173, 2007.

With respect to computerised IS, those used in the 1950s by government and the private sector were primitive in comparison to those in use today [1]. They were designed by technologists to be used by teams of technologists and their support staff. A high level of skills and knowledge was needed to operate these IS/IT systems and a system had to be closely monitored during operation. Today the average users are no longer technologically oriented; instead, they can be any person who understands how to manage an actual IS [2].

2 Purpose and Aim

This article give an introduction to Human Factors relate to Public Information Systems. The purpose of this article is twofold: Firstly to introduce to the synopsis of Human Factors and find practical implication in design of Electronic Public Information Systems (e-PIS).

3 Human Factors an Interrelated Concept

The fundamental philosophy of HF is that all products, equipment, and systems are ultimately made for people and should reflect the goals of user satisfaction, safety, and usability.

The field of Human Factors is the scientific discipline that attempts to find the best ways to design products, equipment, and systems so that people are maximal productive, satisfied, and safe, high interrelationship with Human Computer Interaction.

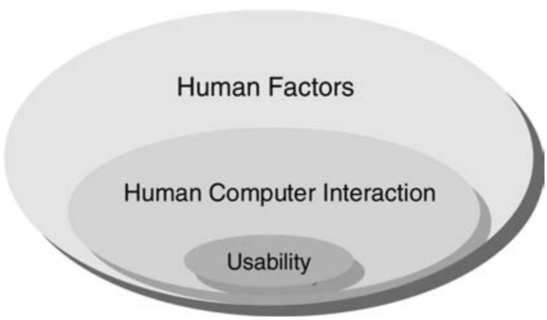


Fig. 1. Interrelationship of concepts in human factors [3]

In frame of citizens/customers use of e-PIS the dominant HF in design are frequently considered the capabilities and limitations of the user from physical, physiological, sensory to psychological factors, in Table 1.

Perception is the ability to detect, identify and recognise sensory input. Perceptual characteristics are important in the design and arrangement of displays and information presentation.

Table 1. Human factors associated to individual characteristics [3]

HUMAN FACTORS	INDIVIDUAL CHARACTERISTICS
Physical Factors	Individual physical dimensions, body posture, repetitive motion, physical interface
Physiological Factors	Human reactions to environments, strength (lifts, grip, carrying, etc.), endurance
Sensory Factors	Hearing, vision, touch, balance
Psychological Factors	Human needs, attitudes, expectations, motivations Mental ability, skills, decision making, training requirements

Cognition refers to higher level mental phenomena such as memory, information processing, use of rules and strategies, hypothesis formation, and problem solving. Therefore, a designer should develop easy-to-use retrieval systems, taking advantage of well-established semantic and symbolic techniques for screen and menu design. Designs consistent with ingrained habits will facilitate performance and reduce training time. Designs that conflict with such habits can lead to errors.

Norman [4] was confused to understand why engineers made things difficult to use, and he suggested applying psychology to engineering to create more useable things. He found that people should be able to comprehend immediately how to use simple things such as ATM machines without the need for instructive signs. That is, by building on the concepts of psychology, things should map to human thought processes in terms of mental models. Mental models, based on experience, people form abstract concepts about how complex information systems actually perform. This is a multifaceted issue, because individuals differ in how they mentally integrate and conceptualize the system.

User attitudes tend to be against change when new technology is seen as bringing desired, particular; acceptance is much more likely when computing is presented as matching to human skills, enhancing rather than replacing them [5], [6].

3.1 Human-Computer Interaction (HCI)

By definition of HCI is discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them [7]. That's means, HCI is the study of how people design, implement, and use interactive computer systems, and of how these systems affect individuals, organizations, and society [8], [9]. Thus, HCI deals with people and computers and the ways they influence each other:

- The design and use of devices such as a mouse, trackball, touch screen, hand-held device,
- The impact of a screen design on users' efficiency and effectiveness during application use.
- Techniques for data representation to improve decision making,

- How to design and evaluate applications or systems that support groups of people,
- Navigational aids for moving around a system,
- Social issues in computing,
- Cognitive modelling of users (mental models).

3.2 Usability

Usability is defined as a high level quality objective: to achieve effectiveness, efficiency and satisfaction, addresses specific issues of human performance during computer interactions within a particular context [10], [11], [12].

Usability gives also many benefits that can include "increased productivity, enhanced quality of work, improved user satisfaction, reductions in support and training costs and improved user satisfaction" [13].

The role of User Centred Design (UCD) in a development effort is to deal with usability and user needs in a way that highlights current difficulties and ensuring that products, systems, and services will [14], [15], [13].

Usability laboratories have become an important tool when prototyping new IS. These laboratories, together with appropriate "field kit" evaluations, enable us to meet the growing number of requests for usability support, chiefly user based evaluations of development prototypes and competitors' products.

4 Identification of Human Factors Problem in Public Information Systems

A adequate short definition of Public Information Systems (PIS) are information systems for public use, and the purpose of a PIS is to provide some kind of service or support to a public process, or process involving "the general public" or "society at large". In contrast, a "non-public" or private information system provides services to some rather specific users closely associated with a certain organisation, performing some specific tasks that are often internal to the organisation [16].

Public information systems are rapidly increasing in variety, size, complexity, and sophistication. Depending on ones definitions, PIS can range from those running on a standalone platform, e.g., a terminal device, to those involving world-wide networks, distributed database, and enterprise-wide interoperability.

Regardless, some common elements among this vast range of systems are that they have human safety implications and they have interfaces with human users. The "users" of a given interface can be citizens, customer, clients, or system operators, administrators, or developers. Of course, e-PIS in the further need multiple channel and interfaces solutions.

Which user characteristics to assess for what tasks are also at issue when determining the needs of a particular system under development; a data set applicable to every design problem has yet to be identified [17]. In other words, to the despair of system design engineers, it is not possible to refer to a table of human tolerances in the

same way it is possible to refer to a table of electronic component performance parameters. On less technical levels the absence of HCI in design is also an important factor in the development of products that are accepted by typical users. For example, when evaluated from an HCI perspective, many simple improvements to a commercial Internet site can be identified [18].

In addition to performance and preference differences between individuals there are wider considerations such as the influence of culture. For example, cultural factors may cause symbols to have different meanings or interpretations in different parts of the world (e.g. [19], [20]. An e-PIS that uses Northern European HCI parameters may not as usable in Asia.

When assessing systems from a human factors perspective, the key difference between public and private appears to be system accessibility.

In one study of Sandberg and Sundberg [21] identify emergent practices in the development of public administration and e-services among four Swedish nationwide government departments. Various governmental Internet sites were analysed to gain a broad understanding of contents and services available at the time of the study. The results shown very clearly that human factors are one of the key problems and need to be solved by system designers of PIS:

- Technology - Unresolved technological problems remain. Software reliability, hardware and software obsolescence, inter-interface operability, quicker and simpler access channels to information, clear transaction requirements, access to humans in the event of problems are some of the issues that can make e-government less than it appears to be.
- Organisational blurring - Organisations change as they implement and depend more upon new technologies; these changes are internal and external. Internally, hierarchies are disturbed, boundaries are blurred, learning processes re-allocate statuses and merits, interactions become more efficient and more complex. Externally, new avenues for networking are opened but with transactional and legal statuses which need to be purposefully considered.
- Security - Information security issues have not been solved with existing technologies and may need yet undeveloped technologies.
- Customer Relation Management (CRM) and new channel strategies to access customers, market and public administration systems.
- Assessment - Accurate ways to assess e-government performance are not being routinely built into new systems.
- Cost - Cost assessment of e-governance projects is not comprehensive. Hidden costs, external costs, transaction costs and reorganisation/learning costs are not well understood. Government is proceeding with implementation without having a clear understanding of cost. The advantages that e-government appears to offer seems to cause administrators to be blind to hidden costs.
- The "Digital Divide" - The problem that a significant percent of the population is usable to access e-government is not adequately addressed. Changing technology and systems compound the problem as those marginalized persons who do learn one system may "lose" access when there is change.

5 Applying Human Factors on Design of Electronic Public Information Systems

Including HF factors in the earliest stages of any design process ensures that a product is defined in terms of user needs [22]. The HF input at these stages also helps project management to redirect resources to approaches that are most likely to have the greatest user benefit. The HF specialist involvement throughout design processes results in catching problems early on; thus helping to reduce overall project costs. Finally, the ability to anticipate user responses continues to improve as successive prototypes more closely approximate the final product and actual field conditions. The HF specialist must be able to base user interface designs and system requirements on an understanding of user needs, goals, characteristics, and expectations. The tasks of systems include translating stated needs into specific technical design requirements that may be implemented within cost targets. The HF specialist can work with public sector processes to balance usability requirements with user constraints, system requirements, and development costs [23]. With this approach, the result can be that the final product functions as desired and in ways that are easily usable; in this instance easy to use by government employees and the general public who wish to access government through IT systems.

To understand the basics implicated in HF on design of e-PISs, a theoretical framework is presented in Figure 2. Its major fundamentals are users (e.g., citizens, and customer), computers, context or environment, tasks, information, interactions, and time [3].

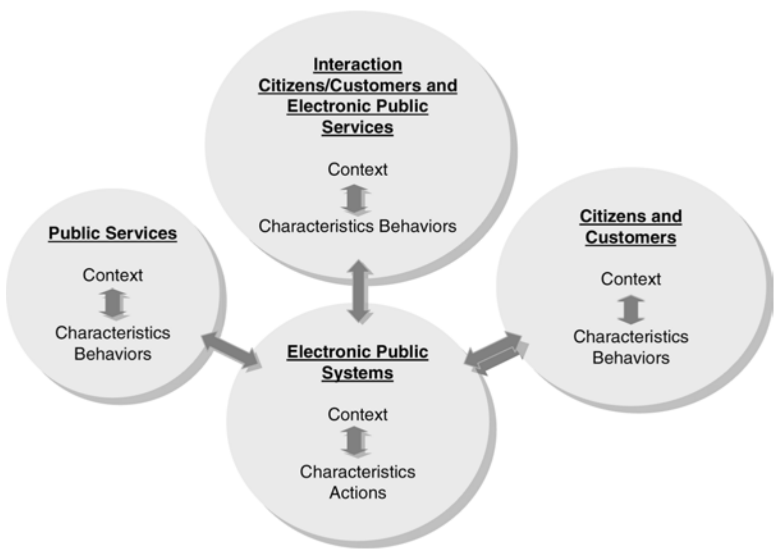


Fig. 2. Framework for Human Factors in Electronic Public Information Systems

Citizen-e-PIS interactions consist of managing and communicating information in a context. During this information exchange, suppliers, citizen, and computer performance occur relative to their respective characteristics. For suppliers (e.g. governments, local authorities), this may be their experience with e-PIS or their formal education. Interactions occur within a context, even a virtual one, and the interactions develop as they move across time. The flow of information is as follows.

Humans initiate the process by turning on a system and sending information to the computer interface. From the interface, the request for information is processed through the computer according to its characteristics.

Use of the framework elements and the application of HF concepts to e-PIS imply a deep understanding about the nature of the work of the suppliers; their context or environment; the way they behave and think about work, context, and tasks; and the tasks themselves.

5.1 Planning for Human Factors

Electronic Public Information systems' planning is a common problem for organizations, and planning for HF is a particular problem within the overall problem of design [24]. The traditional approach to planning IT systems is called the Systems Development Life Cycle (SDLC) method [25]. This approach can be summarized as a linear checklist of activities undertaken in a series of phases that begins with planning and ends with system termination. To put it another way, SDLC is a planning process that separates an IT system problem into its constituent parts, treats each step separately and then unifies the parts into a whole at the end [26].

Prototyping in HCI development work is the process of creating quickly all or part of a system in the form of a simulation. The advantage of prototyping is that it allows user feedback early enough to accommodate major structural changes before extensive development investments are committed. Prototyping is seen as a team process overcoming SDLC. This SDLC model is now seen as being too linear, lacking a group planning orientation and not an actual reflection real world conditions.

5.2 Implementation Issues

Although improved technology must always be considered as a strategy it is often the case that human factors considerations will take up the majority of the time of IT designers and managers. Much is known about the nature of human error, the conditions that encourage error and software designs that are error resistant [4], and much of this is now well recognized in IT training literature [27].

5.2.1 Cognitive Aspects of Implementation

Overcoming employee resistance is a problem in cognitive change. It was found that computerization alters the flow and content of information, changes job content, and affects relationships between organizational members. A key finding was that the primary issue was not computing itself, but rather perceptions of the method of implementing it. It is precisely in this area that human factors become important.

In IT implementation there are common cognitive perceptions that overcome among users: 1) that automation is being mandated arbitrarily, 2) that the new information

system will be unreliable; 3) that the new system will increase rather than decrease the work burden; and 4) that users will neither understand a system nor be able to operate it. Free and open communications throughout an e-PIS implementation process is an important avenue for reducing employee resistance coming from these issues. Negative perceptions regarding the introduction of e-PIS may be overcome to a significant extent through the encouragement of employee decision making during an IS/IT introduction process.

5.2.2 Attitudes

Although user attitudes tend to be against change when new technology is seen as bringing desired benefits attitudes will adapt. In particular, acceptance is much more likely when computing is presented as complementary to human skins, enhancing rather than replacing them [5], [6].

6 Discussion and Conclusion

This article provides an overview of human factors, especially as they relate to Public Information Systems. A rationale is given for the utility of these concepts for advanced design and implementation of future e-PIS.

Further study can provide useful information to scholars and practitioners in the e-PIS and the traditional HF fields for future research, collaboration, publication, and practice.

The fast development and pervasive use of ICT prompt a need to re-examine the broad HF issues in light of the e-PIS development, actual use, and impact on all aspects of our lives. The new framework proposed in this paper emphasizes the entire interaction cycle between humans and technology, rather than a stage or part of it. It also brings in the tasks and contextual factors. This view is intended to show the dynamic as well as the evolutional aspect of issues and concerns regarding the interactions between humans and technology.

The assessment of a limited collection of HF studies, a wide range of research issues and topics being studied by PIS researchers over the last years. The dominating issues fall in the area of e-PIS use, evaluation and impact. PIS researchers are more concerned with issues that occur alter e-PIS is developed. These concerns are closely related to humans' perceptions, beliefs, behaviour, attitude, satisfaction, performance and productivity, and individual differences. Among the small percentage of studies focusing on the development stage, HF researchers are concerned with user involvement and participation, user analyst differences and interaction, programmer cognition studies, and design methods.

References

1. Grudin, J.: The Three Faces of Human-Computer Interaction. IEEE Annuals of the History of Computing 27(4), 46–62 (2005)
2. Casey, S.M.: Set Phasers on Stun: And Other True Tales of Design, Technology, and Human Error, 2nd edn. Aegean, Santa Barbara, California (1998)

3. Staggers, N.: Human Factors. Imperative Concepts for Information Systems in Critical Care. AACN Clinical Issues 14(3), 310–319 (2003)
4. Norman, D.: The design of everyday things. Doubleday, New York (1990)
5. Rosenbrock, H.H.: Engineers and work that people do. IEEE Control Systems Magazine 1(3), 4–8 (1981)
6. Petheram, B.: An approach to integrating technology in human service applications. Computers in Human Services 5(1-2), 187–195 (1989)
7. Hewett, T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G., Verplank, W.: ACM SIGCHI Curricula for Human-Computer Interaction. chapter 2: Human-Computer Interaction (1994) (Rev. 2004-06-03). Retrieved February 19, 2006, from http://www.sigchi.org/cdg/cdg2.html
8. Aspden, P., Corrigan, J.M., Wolcott, J., Erickson, S.M.: Patient safety: achieving a new standard of care. National Academies Press, Washington, DC (2004)
9. Bogner, M.S.: Human error in medicine. Lawrence Erlbaum, Hillsdale, NJ (1994)
10. Rubin, J.: Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. John Wiley, New York (1994)
11. Kushniruk, A., Patel, V., Cimino, J.: Usability testing in medical informatics: cognitive approaches to evaluation of information and user interfaces. In: Proceedings of the American Medical Informatics Association Annual Symposium, pp. 218–222. Hanley& Belfus, Philadelphia, Pa (1997)
12. Lin, H., Choong, Y., Salvendy, G.A.: Proposed index of usability: a method for comparing the relative usability of different software systems. Behaviour & Information Technology 16, 267–278 (1997)
13. ISO 13407. Human-centred Design Processes for Interactive Systems. In: International Organization for Standardization, Geneve, Switzerland (1999)
14. Preece, J., Rogers, Y., Sharp, H.: Interaction Design. Wiley, Chichester (2002)
15. Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T.: Human Computer Interaction. Addison-Wesley, Harlow (1994)
16. Sundberg, B.: What is a Public Information Systems? International Journal of Public Information Systems, vol. 2005:(1) (2005)
17. Dillon, A., Watson, C.: User analysis in HCI. International Journal of Human-Computer Studies 45(6), 619–637 (1996)
18. Benbunan-Fich, R.: Using protocol analysis to evaluate the usability of a commercial web site. Information & Management 39, 151–163 (2001)
19. Syarief, A., Giard, J.R., Detrie T. & McBeath M.K.: An Initial Cross-Cultural Survey of User Perception on Web Icon Design for Travel Websites. In: Paper from the 6th Asia Design International Conference. (2003) Retrieved February 16, 2005 from http://www.idemployee.id.tue.nl/g.w.m.rauterberg/conferences/CD_doNotOpen/ADC/finalpaper/013.pdf
20. Piamonte, D.P.T.: Using multiple performance parameters in testing small graphical symbols. Doctorial Dissertation (2000:02), Luleå University of Technology, Luleå, Sweden. (1999) Accessed February 16, 2006 from http://epublitith.se/1402-1544/2000/02/LTU-DT-0002-SE.pdf
21. Sandberg, K.W., Sundberg, H.: E-goverment: An improvement in service? IADIS International Conference, e-Society 2004, Avila, Spain (July 16-19, 2004)
22. Bailey, R.W.: Human Pelformance Engineering, 3rd edn. Prentice Hall, Upper Saddle River, NJ (1996)
23. Helander, M.: Handbook of Human-Computer Interaction. Elsevier Science Publishers B.V, Amsterdam. The Netherlands (1997)

24. Kuark, J.K.: Factors in the organizational and technical development of CIM strategies and CIM systems. Doctoral dissertation, Eidgenoessische Technische Hochschule, Zurich (1996)
25. Davis, G.B., Olsen, N.: Management Information Systems. McGraw-Hill, New York (1985)
26. Langefors, B.: Theoretical Analysis of Information Systems. Lund University, Lund, Sweden (1973)
27. Salvendy, G.: Handbook of Human Factors. John Wiley and Sons, New York (1997)

Defining a Work Support and Training Tool for Automation Design Engineers

Paula Savioja[1], Leena Salo[1], Outi Laitinen[2], David Hästbacka[2], Topi Judén[2], and Ville Valve[2]

[1] Technical Research Centre of Finland, VTT
Paula.Savioja@vtt.fi, Leena.Salo@vtt.fi
[2] Tampere University of Technology
Outi.Laitinen@tut.fi, David.Hastbacka@tut.fi

Abstract. This paper introduces a work support and training tool (Autaki) which was developed to support automation design engineers in sharing knowledge and learning at work. The development of the application was preceded by an explorative study of automation design engineers' work. Common problems, information needs, learning, and competence factors in the work were studied with an activity-theoretical approach called core-task analysis. The application is an interactive learning environment based on web technologies and emerging standards, such as SCORM (Sharable Content Object Reference Model). Both the core-task of automation design engineers and the application will be discussed.

1 Introduction

Automation engineering is an engineering discipline the objective of which is to design and implement the automated functioning of industrial processes. The output of automation engineering is an automation system which governs the functioning of an industrial process on various levels, ranging from basic process control to higher level processes such as management of production etc. Automation design is carried out in projects in which a particular automation product is used. The designers gather information about the required functionality from different stakeholders. Typically in a large investment project, e.g. construction of a new process plant such as paper mill or power plant, automation covers about 10% of the overall costs of the project. While automation might not seem very significant part of the project, also contradictory arguments can be made. The correct functioning of the automation is extremely vital for the success of the whole investment. Often automation is not only responsible for reaching production goals but also for the safety of process equipment and plant personnel. Malfunctions in automation and errors in its design might lead to severe consequences. Thus, the value of automation design for the success of the whole investment project can be considered to be more than just its share of the overall costs.

Current trends in the industry include making production processes more flexible, safe, and e.g. demand-controlled. These changes increase the complexity of the processes controlled by automation systems and set new kinds of demands for

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 174–183, 2007.

e.g. adaptability on the systems. Thus, also new kinds of competencies, knowledge, as well as work practices and methods, are expected from automation engineers.

The first aim of our study was to find out what these new requirements on automation engineering work are. We also wanted to find out what constitutes competence in automation design and how professional skills are constructed in actual work when the formal education is finished. Secondly, the aim was to discover possibilities to create a work support and training tool to suit the requirements. Thus, the research work presented in this paper both explores the work of automation designers and introduces the concept of a work support and training tool. The purpose of the tool is to help designers cope with the demands of the work and to support the development and sharing of expertise among automation engineers having varying responsibilities and experience levels.

To our knowledge, automation engineering work as such has not been studied previously. However, other types of engineering design work and the competencies required have been studied to some extent. Often the participants in the studies are engineering students carrying out different design tasks. Using students as participants might lead to discoveries and conclusions about development of formal engineering education but the development and construction of professional skills that takes place in the work calls for studies in which professional designers are participants.

Ahmed and Wallace (2004) have compared the knowledge needs of novice and experienced designers. They found out that there were differences in the information needs between the groups. The novices asked a lot of questions about e.g. vocabulary: “what does this mean? How does this or that equipment work?” whereas the more experienced designers posed more complicated questions. From our point of view this means that the work support and training tool should be somehow adjustable to the experience level of the user. According to the study of Ahmed & Wallace, only in 35% of the cases the novices were able to correctly formulate the question. This indicates that for our work support application a simple question and answer tool is not enough. It might be the case that the inexperienced automation engineers cannot even formulate the question to ask from the more experienced engineers when they meet problems in their work.

Development of good work practice is what we want to enhance with our work support and training tool. We share the view of Adams et al. (2003) that good practice can be characterised as reflective. With reflective work practice we mean that an individual is constantly attuned to noticing whether s/he is acting according to the requirements of the work. That means that the person must relate his/her work practice to the objective of the work. Adams et al. found out that the outcome of engineering design work is better when the work practice is reflective. Thus, the question is how we can support the development of reflective practices with an information system. This should be considered when requirements for the work support and training tool are created.

Sandberg (2000) has studied the work of automobile engine optimizers with an interpretative approach. He concludes that the optimizers’ conceptions of work constitute their competence. The broader the conception of work was, the more competent the optimizers were. As we also want to know what constitutes competence in modern automation engineering we decided to study the engineers’ conceptions of their work.

2 Research Methods

The research approach applied in the study is called core-task analysis (CTA). The methodology has been developed at VTT (Technical Research Centre of Finland) to study work in complex, dynamic, and uncertain environments (see e.g. Norros 2004). Previously core-task analysis has been used to explore the works such as nuclear power plant operation (e.g. Savioja & Norros 2004), clinical anaesthesia (Norros & Klemola 2005), and ship manoeuvring (Nuutinen & Norros, in press). Thus, in our study CTA was applied for a quite different type of work. Previously all the studies have been conducted in work environments in which the employees control dynamic processes. We wanted to know whether the analysis approach could be applied for design work, and whether valuable information for software development could be elicited with the approach.

The aim of CTA is to identify the *core task* of a particular work. Core task is the essence of work which stays the same from one situation to another independent of e.g. the organisation of the work. The conception of the core task that the different people possess is interesting, because it portrays what they consider *the meaning* in the work. In a well functioning work system the meaning is understood (and the conception of core task shared) by the whole community. In core-task analysis the aim is to find out the varying conceptions of the core task that people have, and to construct an “objective” model of the core task. For the model, the activity system approach developed by Engeström (see e.g. 1999) is used. By comparing the different conceptions of the people, and including in the analysis also the activity system model, it is possible to notice discrepancies and contradictions that lead to problems in practical daily work. The problems we hope to be at least partly solved by the work support and training tool to be developed.

2.1 Collection of Data

Data about automation design work was collected with two main methods: workshops and semi-structured interviews. In the workshops the participants were management level personnel and the topics discussed related to automation design work in general, for example the objectives, organisation, and tools. Altogether three workshops were organised in which also the researchers participated. The interviewees in the semi-structured interviews were regular design engineers. Altogether 8, both novice and experienced designers participated. The interview themes covered the objective of work, the construction of the activity, the criteria by which the output of work is evaluated, the factors that influence the success of work such as tools, information, collaboration, and work practices, the skills and knowledge possessed by a competent automation engineer, and the possibilities and obstacles for learning.

Two Finnish automation design companies participated in the study. Both companies can be characterised to fit Petre’s (2004) norm of successful innovative design companies. They work with multidisciplinary small project teams, are in close contact with research institutes, and are familiar with the technology front end.

2.2 Analyses of the Data

The results of the workshops were saved in workshop memos. The interviews were audio recorded and partly transcribed and turned into interview logs. All the logs were collected to a single spreadsheet so that the answers from individual interviewees could be compared and classified. We tried to discover what the automation engineers considered as the objective in their work, and how they defined a good automation design engineer works. We approached the issues by analysing the answers to the questions concerning content of work, success criteria of work etc. We also looked for variations in the answers of designers. The activity system model was constructed by combining the interview data and the workshop memos. The model was validated in an additional workshop with the participating companies. The understanding of the work acquired by the researchers during the process was used to derive both functional and non-functional requirements for the work support and training tool.

3 Results of the Study

Results of the work conducted are twofold. First of all, there are the results of the interview study. This includes the core task in automation design work, and understanding of how learning and construction of professional skills takes place within the work. The other part of the results constitute the work support and training tool, the Autaki system that has been developed with the knowledge gathered about automation design work. First, a concept for the tool was developed without taking the technical constraints into account. Then a prototype was implemented using particular technologies.

3.1 The Core Task in Automation Design

The model of the core task is defined by analysing the objective of the work and the means that the personnel have to fulfil the objectives. The activity system model is constructed with the knowledge of the afore-mentioned. Central concepts in the model are the *objective* of the work, the *object* of work, and the *instruments* used to operate on the object. The objective of the work reflects the relationship between the object of the work and the outcome of work.

The activity system model of automation engineering work (Fig. 1) portrays the part of the results of the core-task analysis. We found out that the design engineers do not fully share the conception of what is the objective in the engineering design. There are design engineers who see the satisfied customer as successful output. Some of the other engineers emphasize that also the application must work correctly in all conditions despite what the customer might consider good enough functioning. In addition, some engineers extend the objective to cover the development of own work methods and techniques and also the fact that the code must not only work but also be legible and understandable by other engineers. There were also different conceptions of the object of work, as is answers to the question “What happens in automation engineering?”. The answers varied from “automation of individual equipment” and “changing information from one format to another” to “implementing correct

functioning of an industrial plant”. This clearly shows that some automation engineers think of the role of automation as more vital in the whole engineering project than the others.

The differences in the conceptions can probably be explained by the following factors:

- The educational background
- Years of experience in the work
- Type of work carried out during the career so far
- Factors related to work culture

In real world there are often contradictory objectives in the work that create tension and lay demands on the workers. For example, safety and cost efficiency are often both considered objectives of process control work (in addition to actual production). This means that work practices which help people to cope with the demands must be created and maintained. CTA aims also at recognising what these good work practices are.

As recognising work practices, we mainly concentrated on the design engineers’ practices when facing difficulties or problems in the work. Some turn to colleagues to get help; the others use literal sources to find answers to questions (this was quite rare though). Often in a project there is a person who is experienced and whom many other engineers use as an information source in problem situations. Especially the inexperienced designers claimed that they wanted to carry out some pre-research before turning to this person for answers as they wanted not to “waste his valuable time”.

When asked to describe their professional learning experiences, many design engineers said that the implementation of the application in the field, at the real plant, was the most important learning experience. The problem is that often the people who would need this learning opportunity, such as the inexperienced designers, often cannot take part in the field work due to financial or schedule reasons. It is also a problem from educational point of view, that other people take responsibility of the application after it is implemented in the plant. This means that the design engineers do not get direct feedback for their work. They seldom get to know how well their individual solutions function in the plant, and this way reflection on their own work is partly inhibited.

Automation engineering as a discipline undergoes relative turmoil because of the constantly developing information technology. This creates learning requirements for all the engineers independently of previous experience.

One of the characteristics of the automation engineering work is that it is carried out in an engineering middle field. This means that the automation engineers have to gather information from all the other fields of engineering, e.g. process, chemical, piping, and electrical; and with this knowledge interpret the intended functioning of the plant. This calls for good co-operation and communication skills from automation engineers. They also have to know the “languages” of the other disciplines in addition to their own to avoid misunderstandings.

Often the automation engineer has to start designing the automation before the other disciplines are ready with e.g. process design. Sometimes even the customer’s requirements are not yet complete. This means that the automation engineer must partly work with uncertain information. When asked about the characteristics of a good automation design engineer a few of the interviewees answered that one has to know what can already be decided about automation – even though the requirements for automation are not complete.

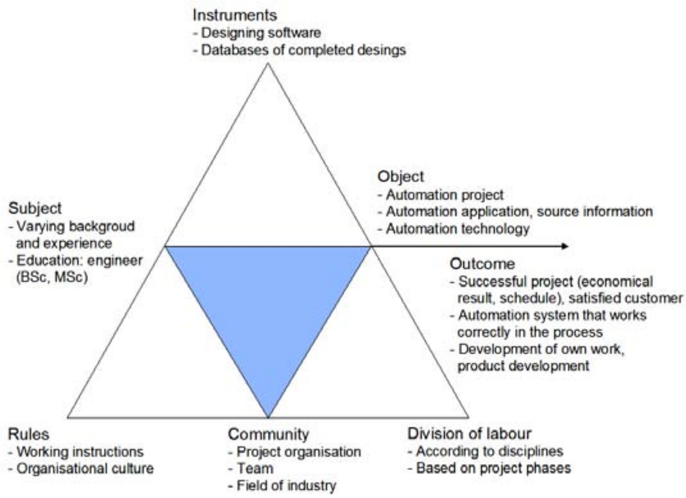


Fig. 1. Activity system model of automation design engineering work. The model depicts the varying conceptions of the interviewees.

Core-task analysis aims at recognising the core content of a particular work. The core task is something that prevails from one project to another and brings consistency to the job. A core task can be expressed as core-task demands, the coping with which signifies success in the work. Core-task demands identified in this study to characterise the work of automation design engineers can be summarised as:

- Adaptation to the technology change, maintaining and gaining technical skills
- Co-operation and communication with different engineering disciplines
- Working with uncertain information

As we were only able to interview eight automation design engineers the results can only be considered explorative. We are not able to explain the differences in the conceptions of the engineers by any comparative means. In addition, we can not say that we have covered the whole field of automation engineering, and thus, found all the possible varying conceptions. However, we had enough information to start the requirements definition for the work support and learning tool.

3.2 Concept for the Autaki Application

The starting point for the development of the Autaki concept was the definition of work demands identified in the core task analysis. It was concluded that the tool should provide automation engineers means for maintaining and gaining technical skills and knowledge, support co-operation and communication, and promote sharing of knowledge related to e.g. good working practices to cope with uncertain information and other challenges in work. These general requirements were considered more thoroughly from the point of view of different groups of users and situations of use. Descriptions of scenarios depicting how, when, and why particular users would want to use the tool were written.

The scenarios depicted e.g. the following usage situations. In the first situation an inexperienced automation design engineer searches for an answer to a technical question considering process measurements. In another situation the same designer wants to comment on a design template while s/he is implementing an automation application with programming software. When the engineer has more time in use s/he can go more deeply into a new subject s/he is not yet familiar with. In a fourth situation an experienced design engineer wants to share his/her knowledge on project life cycle, e.g. which tasks can be conducted in a situation when part of source information is missing. For each scenario required information and functionalities were specified.

When building the concept for the Autaki application it was defined that in order to serve the diverging information needs of different user groups the tool should provide information at least of the following topics: examples of good design solutions, descriptions of different industrial processes and process components, information of the phases and tasks in an engineering project, and latest information of new techniques and trends in the field. As the number of various information sources is high at the moment and there are problems in finding information, the application should function as a unified user interface for various databases and information sources. For supporting information finding the application should include advanced search tools and various ways for navigating in the content. Different views to the content should be provided, i.e. the content should be structured and presented to the users in several ways, for example as manuals or textbooks. The idea of views is to take into account different situations of usage in which the way of using information varies.

To promote individual learning the contents of the application should be arranged to courses which the users can take whenever they have time for studying. Also, to support managing the contents, the tool should provide means for making personal notes and bookmarks. According to the interviews, development of expertise often takes place through guidance by more experienced design engineers. Therefore collaborative features promoting sharing of knowledge in the community should be included in the application, for instance, question and answer forums and possibility to send email to experts on the field. The users should be able to add comments to all contents accessible through the tool, including project documents, design templates, etc. To bring the information easily available while working with other tools, the application should be integrated to designing software. The main requirements elicited for the work support and training tool are summarised below:

- The application should support daily work of engineers by providing useful information taking into account the diverging information needs of different user groups
- The application should support both collaborative and independent learning at work as well as knowledge sharing by providing means for e.g. commenting, asking and answering questions, etc.
- The application should provide different views to the content e.g. manuals and courses as well as advanced search tools and multiple ways of navigation
- The application should provide means for editing, adding, and removing material to keep the contents up-to-date
- The application should provide as means for creating personal bookmarks and notes
- The application should be integrated into existing tools and practices

From the technical point of view, the Autaki prototype developed during this research project fulfils essential parts of the Autaki concept described above. Prototype consists of two core parts (see Figure 2).

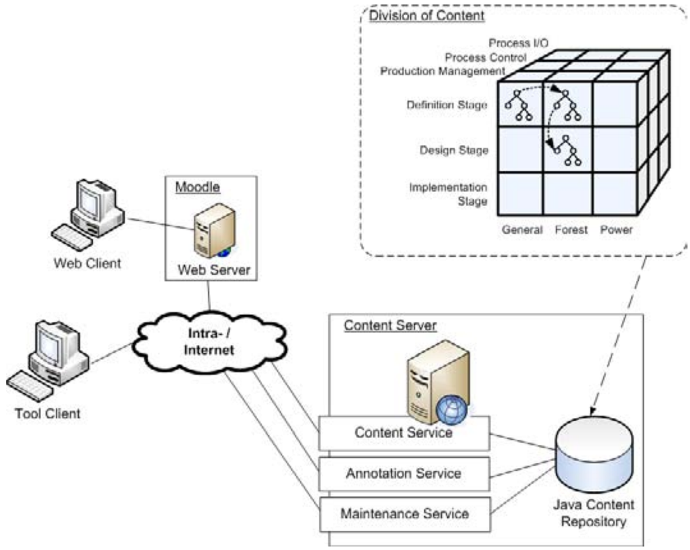


Fig. 2. Autaki architecture and the structure of the content

The content server can be split into a content repository and content services. All Autaki materials are located in the content repository which is an open source Java Content Repository (JCR) implementation: Apache Jackrabbit (see e.g. Sommers 2005). The content consists of material modules, and the modules consist of several material components. These modules and components are built up according to SCORM (Sharable Content Object Reference Model) specification (see e.g. ADL, 2004) using IMS Global Learning Consortium specification for Learning Object Metadata (LOM). Industry specific terminology improves the accessibility of the information Autaki contains. The idea of the industry specific terminology is introduced in Figure 2 in the Division of Content. Three dimensions are used to structure the content:

- Lifecycle (definition stage, design stage, implementation stage…)
- Automation hierarchy level (production management, process control, process I/O)
- Industry (general, forest, power…)

The content server includes all content services it needs to execute the requests sent by a client. Services are implemented as web services (see e.g. Alonso et. al., 2004). The concept of Autaki defines three kinds of services: *Content Service* deals with requests related to data handling (e.g. search), *Annotation Service* deals with annotation related requests (for commenting content, asking questions etc.), and *Maintenance Service* handles all tasks of maintaining content in the content repository. In the

prototype only two of the former services were implemented. Each web service offers WSDL-based (Web Services Description Language, see e.g. Alonso et. al. 2004) interface, which makes the services easier to access via the Internet. Communication between the content server and clients is based on SOAP (see e.g. Alonso et. al. 2004) which respectively is based on XML messages (Extensible Markup Language, see e.g. W3C 2004).

Services introduced above can be used by various clients. For instance, in the prototype of Autaki the client is either a web client application or a tool client application. The web client is the main user interface of Autaki, and it includes all needed functions. The designer uses the web client from his own workstation using a web browser. The tool client, on the other hand, is an optional client type of Autaki application. The idea behind the tool client is to bring the work support and training tool as close to designers as possible by integrating it into engineering software tools. During this development project a tool client was implemented as a demonstration of the possibility.

The whole Autaki prototype is implemented utilising open source technologies and tools. For instance, the open source web server software Axis2 (see e.g. Perera 2005) used in the content server and the web-client application is based on the Moodle course management system (see Moodle). Moodle is a broadly used and rapidly progressing web based course management system (CMS). The web client of Autaki is basically an independent extension module for Moodle. The client utilises some features of Moodle such as user authentication and management of sessions but all Autaki specific features are implemented as autonomous applications.

4 Conclusions

In the study described above, the concept of a work support and training tool for automation engineers conducting automation design work was developed based on work analysis. In the work analysis a specific method, the core-task analysis, was used to understand the demands of the job, and thus the needs for work support and training tool.

We were able to identify variation in automation engineers' conceptions of what is important in their job. We also discovered demands that the engineers have to cope with in their daily work.

The developed application is an interactive learning environment based on web technology. The users can enter the Autaki application either through a web client or a tool client which is embedded in the design software.

In the next phases of the research the use of the prototype tool will be examined in more detail.

Acknowledgements. The authors wish to thank the personnel from Metso Automation and Insta Group who have participated in the study. We also wish to thank Professor Seppo Kuikka from Tampere University of Technology and senior research scientist Teemu Tommila from VTT for their valuable insight in the research project.

References

1. Adams, R., Turns, J., Atman, C.: Educating effective engineering designers: the role of reflective practice. Design Studies 24(3) (May 2003)
2. ADL (Advanced Distributed Learning) SCORM – Sharable Content Object Reference Model: SCORM Content Aggregation Model version 1.3.1 (2004)
3. Ahmed, S., Wallace, K.M.: Understanding the knowledge needs of novice designers in aerospace industry. Design studies, 25(2) (March 2004)
4. Alonso, G., Casati, F., Kuno, H., Machiraju, V.: Web Services – Concepts, Architectures and Applications. Springer-Verlag, Heidelberg (2004)
5. Engeström, Y.: Activity theory and individual and social transformation. In: Engeström, Y., Miettinen, R., Punamäki, R.-L. (eds.) Perspectives in Activity Theory, pp. 19–38. Campridge University Press, Campridge (1999)
6. Moodle. Moodle Community's web pages. (Checked 2006-09-10) http://moodle.org/
7. Norros, L., Klemola, U-M.: Naturalistic analysis of anaesthetists clinical practice. In: Montgomery, H., Lipshitz, R., Brehmer, B. (eds.) How experts make decision, pp. 395–407. Lawrence Erlbaum Associates, Hillsdale, NJ (2005)
8. Nuutinen, M., Norros, L.: Core Task Analysis in accident investigation – analysis of maritime accidents in piloting situations. Cognition, Technology & Work. (In Press)
9. Perera, S.: Introducing Axis 2, the Next Generation of the Apache Web Service Stack (2005) Developer.com (web publication: http://www.developer.com/services/article.php /3525481)
10. Petre, M.: How expert engineering teams use disciplines of innovation. Design Stuies 25(5) (September 2004)
11. Sandberg, J.: Understanding human competence at work: an interpretative approach. Academy of Management Journal 43(1), 9–25 (2000)
12. Savioja, P., Norros, L.: Evaluation of operator practices as means of integrated system validation. In: Proceedings of the Fourth American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation. Controls and Human-machine Interface Technologies, Columbus, Ohio (2004)
13. Sommers, F.: Catch Jackrabbit and the Java Content Repository API (2005) Artima Developer (Web publication: http://www.artima.com/lejava/articles/contentrepository.html)
14. W3C 2004, XML Schema Part 0: Primer Second Edition, W3C Recommendation (28 October 2004) Retrieved 22 January 2007 from http://www.w3.org/TR/xmlschema-0/

A Comparative Study of Multimodal Displays for Multirobot Supervisory Control

Boris Trouvain and Christopher M. Schlick

Forschungsgesellschaft für Angewandte Naturwissenschaften,
Neuenahrer Str. 20, 53343 Wachtberg-Werthoven, Germany
Institute of Industrial Engineering and Ergonomics, RWTH Aachen University,
Bergdriesch 27, 52062 Aachen, Germany
Trouvain@fgan.de, c.schlick@iaw.rwth-aachen.de

Abstract. The supervisory control of ground-based mobile multirobot systems requires to perform multiple concurrent tasks under high levels of time pressure resulting in heavy workload. In this paper we present the design and evaluation of multimodal displays for a particular problem associated with the supervisory control of ground-based multirobot systems: the coordination between the platform specific robot control task, e.g. navigation and obstacle avoidance, and the mission specific payload task. The coordination requires the operator to concurrently monitor and switch attention between the robot control and the payload control tasks depending on the mission requirements. Multimodal human-robot interfaces can significantly support human information processing by communicating information across multiple channels and can therefore improve concurrent task processing. An experiment was designed and carried out with 14 participants which compares four human-robot interface configurations with a simulated two-robot ground-based multirobot system. The results show that the multimodal interfaces perform significantly better across multiple variables and have the lowest workload. Based on our gaze tracking results we can conclude that our multimodal interface has an effect on the visual scanning behaviour in the peripheral regions of the camera display.

Keywords: Human-Robot-Interface, Multirobot, Multimodal.

1 Introduction

Human-multirobot systems are a promising approach to tightly integrate human and artificial intelligence in complex missions. The user interface is an important component of such a flexible system. In order to maintain situation awareness the operator must continuously perform the mental fusion of displayed information. Due to the well-known human limitations in concurrent information processing [15] it is important to strive for novel human-multirobot interface designs. The innovative approach of this paper is to design a true multimodal, single screen interface using interface components that allow the shifting of state information of the robots from the visual to the auditory or tactile channel of the operator. Wickens' multiple resource theory [15] predicts that the division of attention across different modalities

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 184–193, 2007.

supports concurrent human information processing and therefore should significantly support operators' mental fusion of information.

Supervisory control of mobile robots can be regarded as a dual task problem: The primary task is the robot control task covering the basic aspects of platform control, e.g. maintaining operational effectiveness, obstacle avoidance, etc. The secondary task is the payload control task which is related to the mission objectives, e.g. to apply sensors for surveillance or search and rescue. However, if one wants to empower a single operator to manage one or two mobile robots, careful consideration of all levels of the human-robot system is necessary. While technological problems of mobile multirobot systems are likely to be solved in the future, the cognitive performance of human operators to effectively control such complex concurrent systems will remain constant. When comparing the technologically feasible levels of concurrency in future multirobot systems and human cognitive performance in "multitasking" there is a strong incompatibility [11].

On the robotic level a sufficient degree of autonomy is a prerequisite for reducing the workload of the operator. Autonomy enables robots to be productive without operator intervention for certain, albeit hardly predictable amounts of time. This particular aspect of autonomy and human-robot interaction is described as neglect tolerance and was extensively studied [3, 10]. For instance, Olsen investigated the relationship between neglect tolerance and effectiveness of human-robot interaction [10]. The interplay between robot autonomy and human-robot interaction was further studied in [5, 9].

A variety of factors for designing effective and efficient human-robot interfaces are introduced by Adams and Goodrich [2, 6]. Although a human-multirobot interface can share and reuse many elements from single robot user interfaces, special attention must be paid to two problems. First, each robot operates within an arbitrary complex context (task, environment etc.) and the contexts are coordinated or independent. In order to effectively supervise the multirobot system the user interface must support the operator in smooth switching between different contexts. Second, a complex set of concurrent tasks must be planned, monitored and modified in order to maintain the system's operational effectiveness.

The problem of how to concurrently work on multiple tasks is a common research question in today's highly automated systems and is investigated in various domains [1, 8]. For instance, findings of Wickens [13, 15] indicate that the visual and auditory system of the human can be regarded as separate information processing resources. This multiple resource approach is particularly interesting for human-robot interaction. Wickens asserts, "… in a heavy visual environment, auditory displays will improve time-sharing performance." and cited 18 studies on dual task performance. The cross-modal and intra-modal approaches were also investigated in a simulation study of the concurrent control of two UAVs [4, 14]. The results revealed positive effects of simultaneous visual and auditory information presentation in subtasks requiring many gaze movements. Helleberg [7] presents a cockpit display study about the compatibility of different types of information with respect to auditory delivery. He found that the auditory-only condition was the least disruptive of ongoing visual tasks.

2 Methodology

2.1 Dual Task Design

The primary goal of this study is to investigate the coordination between the human operator and the autonomous subsystems. To empower the operator when coordinating activities – he or she must decide to either proceed with the mission task or to assist the autonomous subsystems – it is necessary to ergonomically present the required information and to minimize the workload due to attention shifts. Multimodal interfaces allow attention to be divided between, for example, the visual and the auditory channel, and therefore often improve concurrent information processing. In our evaluation we model this situation using a generic dual task design – robot control task and payload control task – as already pointed out in the introduction. The robot control task was an obstacle classification task. In the case of an obstacle that can not be classified by the autonomous navigation system human operator as a supervisory controller has to intervene and interactively assist the autopilot by notifying the autopilot that a water-filled negative obstacle lies in the driving path. The payload control task design is derived from observation and surveillance scenarios which are typical applications of mobile single and multirobot systems. The subjects were informed that both robots of the multirobot system would follow a preprogrammed patrol path. Along this path a number of depressions were placed randomly in parallel to the left or right of the path. These depressions can contain up to two visually hard to detect targets which have to be marked upon detection on screen using the same point-and-click method as in the robot control task. Figure 1 shows an in-simulation screen capture of the experimental setting.

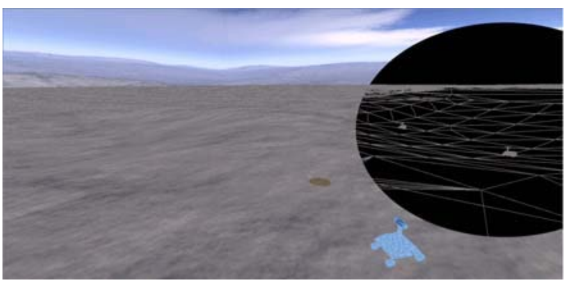


Fig. 1. Screen capture from third-person perspective with superimposed corresponding partial wireframe capture illustrating the task setting. A depression containing two targets can be identified in the wireframe section. Ahead of the robot a water-filled obstacle can be seen.

2.2 User Interface Design

The visual layout of the human-robot interface was inspired by the windscreen-dashboard layout of a car. Since the payload control task in this study is an observation task the upper two thirds of the display space are used for the camera view. The lower third contains the two consoles for two mobile robots (see Figure 2). The design decision to limit the interface to systems with just two robots is based on

previous studies concerning the supervisory control of a multirobot system [12]. A characteristic design feature of the user interface is the mutually exclusive access to the camera view. Therefore, the operator has to actively switch between the robot cameras with a keyboard key in order to get access to the current view. The bottom dashboard contains two consoles for the simulated robots. The graphical layout is guided by horizontal symmetry with respect to the console placement (see figure 2 for example).

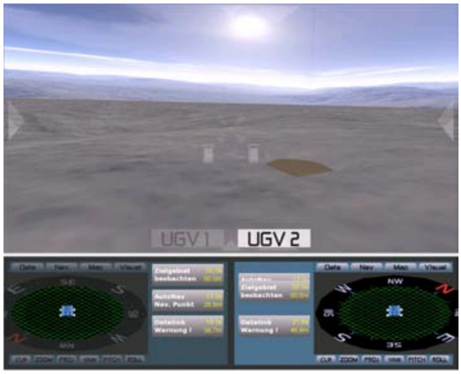


Fig. 2. Screen capture from multirobot user interface as used in this study. The upper section displays the camera view of the supervised robot. The lower section contains two robot consoles; on the left for robot one and on the right for robot two.

2.3 Mono- and Multimodal Display Configurations

This study compares four distinct display configurations, two monomodal and two multimodal, for rendering the information about the remaining time until the obstacle is reached. The graphical design of the first monomodal (that is only visual) user interface was already described in the previous section and simply is a general purpose dynamic priority list. On the contrary, the second mondomodal user interface is a task specific head-up-display allowing the operator to identify its state without interrupting the use of the camera view. The multimodal display configurations are differentiated by their respective channel, auditory or tactile, used to offload information. Figure 3 shows both screen captures and photos of the used interfaces.

Fig. 3. From left to right: The dynamic priority list, the head-up-display, headphone based binaural auditory display and a wrist-watch attached tactile element

Common to all the interface elements is the rigorous left-right structure of information displays following the graphical user interface as shown in Fig. 2. The left priority list belongs to the left robot while the right priority list belongs to the right one. The same is valid for the head-up-display which is rendering the remaining distance to the obstacle by a proportional semi-transparent bar. The head-up-display is positioned in the central region of the camera view below the crosshair.

An important aspect of the auditory display is its binaural design. Humans are well capable of concurrently listening to multiple sound sources and quickly separating the relevant from the irrelevant ones. This effect is used in the context of the design of the human-multirobot interface by delivering the auditory information of each of the two robots exclusively to the left or right ear of the operator. This design allows the operator to concurrently listen to the auditory channels of both robots. Again the same is valid for the tactile interface which is mounted on the left and right wrist of the operator. The signal used for both auditory and tactile information transmission is an interval and pitch modulated beep tone, or vibration, which dynamically encodes the obstacle's temporal distance. The silence intervals between subsequent beep tones are synchronized to play at the same point in time if their interval length is equal.

In order to compare the monomodal and multimodal interfaces the two interface alternatives were parameterized to communicate the same information. Both the auditory, tactile and the visual elements are activated at a temporal obstacle distance of 15 seconds. The vertical movement of the visual element in the pending interaction list is encoded by three easily distinguishable silence interval lengths of the corresponding auditory obstacle signal. When the visual element moves up the pending request list, the silence interval is set to the corresponding obstacle temporal distance level. The temporal distance is the estimated time until the robot hits the obstacle given the current heading and velocity. The three levels were set at 15, 10 and 5 seconds. The corresponding silence interval lengths for the auditiory and tactile interface were 1000 ms, 500 ms and 125 ms (1Hz, 2Hz and 8Hz). The numeric display containing the temporal distance to the obstacle was encoded into the beep tone's, respectively the vibration's, pitch level.

2.4 Design of Experiment

Fourteen subjects with a mean age of 27 years participated in the laboratory study. A written instruction manual containing information about the trials was handed out to the participants several days before the trials. An introductory training session was conducted at least one day before to familiarize the participants with the user interface, the robot and payload control task, and the visual, tactile and auditory displays. Following this training the users conducted four thirteen minute trials using each of the user interfaces. Finally, a post trial subjective workload assessment using the NASA Task Load Index (TLX) was carried out.

The hypothesis for the robot control task is that the multimodal interface improves operator's performance due to the offloading of the task related information to the auditory channel. This should enable operators to improve their coordination between the two concurrent tasks as well as between two robots due to the binaural information presentation.

The hypothesis for the payload control task is that the performance will also increase when using the multimodal interface due to the reduction of visual load, thereby enabling the operator to use more time for the target search in the camera view. Finally, the hypothesis for the workload section is that workload will be lower when working with the multimodal interface due to its capabilities of supporting the coordination between concurrent tasks. We tested the associated null hypotheses using a general linear model (GLM) with repeated measures. The level of significance was $\alpha_{GLM} = 0.05$. The results are shown as error bars with two times the standard error. Variables are named LIS for the priority list, HUD for the head-up-display, AUD for the auditory and TAC for the tactile display.

3 Results and Discussion

Robot control task: Obstacle classification

The variable "obstacle classification" accounts for the correctly classified obstacles. Figure 4 shows the means and error bars of the normalized results for the interface alternatives investigated.

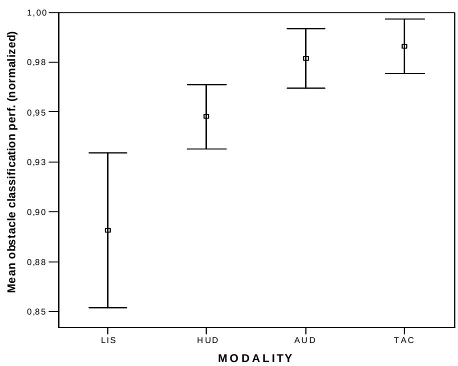


Fig. 4. Mean normalized obstacle classification performance. A total of 55 obstacles were passed by the two robots. The performance of the multimodal interfaces was almost perfect with between zero and two failures.

According to Figure 4 the priority list scores significantly lower than any other interface, especially compared to the head-up-display ($p < 0.01$) representing the best monomodal interface. Both multimodal interfaces outperform both monomodal ones (AUD vs. HUD $p < 0.01$ and TAC vs. HUD $p < 0.01$). The result for the priority list is not unexpected as this interface requires the operator to switch his visual attention frequently between the camera view and the console. The result of the head-up-display (HUD) however is surprising as it is scores significantly better than the priority list but still is also significantly lower than the multimodal interfaces. This indicates that even an optimized interface such as the head-up-display suffers from the conflict with the target search task. The multimodal interfaces show no significant differences.

Payload task: Target detection

The secondary task in this experiment was to detect and mark as many targets as possible while still classifying all incoming obstacles. Figure 5 shows the results for the target detection performance. The priority list again scores significantly lower than any other interface ($p < 0.05$ vs. TAC, $p < 0.01$ vs. AUD, $p < 0.001$ vs. HUD). Given the high potential for conflicts between the visual search task and the necessity to frequently check the list elements this is not unexpected. It is however unexpected that the head-up-display appears to provide the highest mean scores, or the lowest impairment, in respect to the target detection. This is surprising as the target detection task should benefit from the visual offloading provided by the multimodal interfaces. Based on the present data, the performance difference between the head-up-display and the tactile display is significant ($p < 0.05$).

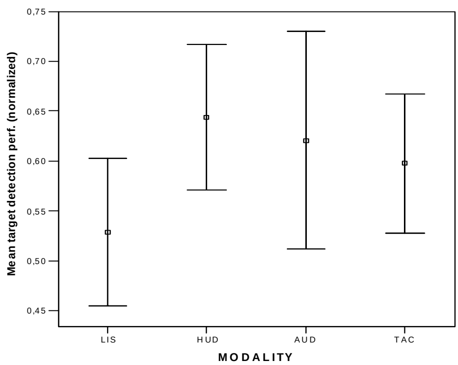


Fig. 5. Mean normalized target detection performance. A total of 59 targets were detectable by the operator.

Gaze Analysis

We used a video based remote eye tracking system (Tobii x50, measuring at 50Hz) to record the visual scanning paths of the participants. As the payload control task is based on a difficult continuous visual search task, we expected to find differences in the eye movements in respect to the modality of the displays. The general visual scanning is focused along the horizon level of the camera view (see Fig. 6 left). Our hypothesis is that the scanning patterns will change especially in the peripheral regions when conflicts with visual perception occur. This means, that the operator can expand the search area. To analyze this we used a set of regions of interest to filter the measured gaze positions along the horizon (see Fig. 6 right).

Figure 6 shows the results of the measured gaze positions in the L1 and R1 region. A clear trend can be identified between the monomodal and multimodal interfaces. The use of the priority list resulted in significantly lower visual activity than with any other interface. The head-up-display requires the operator to often focus the central

crosshair region of the camera view and thus results in a narrower visual scan pattern. Both multimodal interfaces enable significantly higher visual scanning in the peripheral regions than the monomodal ones (AUD vs. HUD $p < 0.01$, TAC vs. HUD $p < 0.05$). This is especially valuable as this can reduce the problem of tunnel vision under high workload. Unexpectedly, the head-up-display shows no significant difference compared to the priority list (HUD vs. LIS $p < 0.1$).

Fig. 6. Gaze analysis design. The left image shows an exemplary hotspot visualization of the gaze activity (fixations). The right figure shows the corresponding regions of interest used as filters.

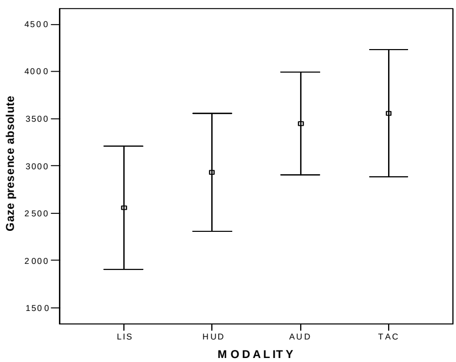


Fig. 7. Number of measured gaze positions in L1 and R1 region (see Figure 6 left)

Workload

Finally, the subjective workload was sampled using the NASA-TLX workload rating technique. Figure 8 depicts the results. Similar to the variables investigated before the priority list based interface does not only provide the lowest performance but requires also significantly higher workload than the other three interfaces (LIS vs. HUD $p < 0.01$). The multimodal interfaces are rated almost identical but do not achieve a significantly lower workload than the head-up-display (HUD vs. TAC $p < 0.25$, HUD vs. AUD $p < 0.25$, HUD vs. AUD $p < 0.95$).

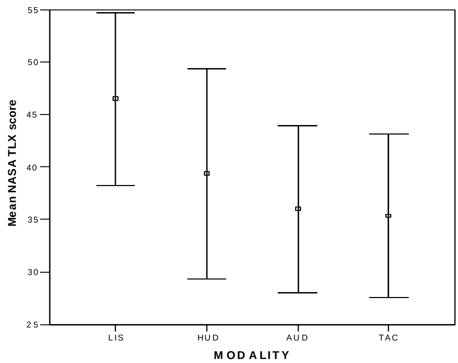


Fig. 8. Results of the NASA-TLX workload sampling. The scale ranges from 0 to 100 where a lower score means lower workload.

4 Conclusion and Future Work

In this paper a study was presented to compare two monomodal and two multimodal alternatives of our experimental multirobot user interface. The results clearly show that the human operator can benefit from our multimodal interface design because the performance of the robot control task is significantly higher when using the multimodal interface (see figure 4). Surprisingly, this performance increase does not have a positive or negative impact on the performance of the payload control task. This is a clear indication that the multimodal user interface adequately supported human multi-tasking. It is even more important that the subjective workload of the operators being measured with the NASA TLX method is significantly reduced in comparison to the priority list interface while the performance increased (see figure 8, 4). It is interesting to see that the results for both the auditory and the tactile multimodal interface perform equally well and therefore can be substituted for each other without reducing performance.

In summary, the initial claim that a multimodal interface is fruitful when striving for significant performance improvements of human supervisory control of multirobot systems is supported by the experimental data of the study. However, the good performance of the monomodal head-up-display demonstrates that multimodal interfaces are not per se superior to monomodal ones.

In future experiments, we will investigate more closely the effect of the modality on the performance of the target search task by improving our gaze analysis. We will further combine the auditory and tactile interface to design multirobot interfaces for systems with up to four robots.

References

1. Adamczyk, P.D., Bailey, B.P.: If Not Now, When? The Effects of Interruption at Different Moments Within Task Execution. In: Proc. of the Conference on Human Factors in Computing Systems, pp. 271–278. Vienna, Austria (2004)
2. Adams, J.A.: Critical Considerations for Human-Robot Interface Development, AAAI Fall Symposium on Human-Robot Interaction (November 2002)

3. Crandall, J.W., Goodrich, M.A., Olsen Jr., D.R., Nielsen, C.W.: Validating Human-Robot Interaction Schemes in Multitasking Environments. IEEE Trans. on Systems, Man, and Cybernetics-Part A: Systems and Humans, 35(4) (July 2005)
4. Dixon, S.R., Wickens, C.D.: Control of Multiple-UAVs: A Workload Analysis, In: 12th International Symposium on Aviation Psychology, Dayton OH (2003)
5. Goodrich, M.A, Olsen Jr, D.R., Crandall, J.W., Palmer, T.J.: Experiments in Adjustable Autonomy, In: Proc. of IJCAI01 Workshop on Autonomy, Delegation, and Control: Interacting with Autonomous Agents
6. Goodrich, M.R., Olsen Jr., D.R.: Seven Principles of Efficient Human Robot Interaction, In: Proc. of the IEEE International Conference on Systems, Man, and Cybernetics, pp. 3943–3948. Washington (October 5-8, 2003)
7. Helleberg, J., Wickens, C.D., Goh, J.: Traffic and Data Link Displays: Auditory? Visual? Or Redundant? A Visual Scanning Analysis, In: 12th International Symposium on Aviation Psychology, Dayton OH (2003)
8. Miller, S.R.: Window of Opportunity: Using the Interruption Lag to Manage Disruption in Complex Tasks, In: Proc. of the Human Factors and Ergnonomics Society 46th Annual Meeting (October 2002)
9. Nielsen, C.W., Goodrich, M.A., Crandall, J.W.: Experiments in Human-Robot Teams, Multi-Robot Systems: From Swarms to Intelligent Automata Volume II (2003)
10. Olsen, D.R., Goodrich, M.A.: Metrics for Evaluating Human-Robot Interactions, In: Proceedings of the Performance Metrics for Intelligent Systems Workshop, Gaithersburg, MD (September 16-18, 2003)
11. Sheridan, T.B.: Telerobotics, Automation, and Human Supervisory Control. The MIT Press, Cambridge (MA) (1992)
12. Trouvain, B., Schlick, C.: A Study of Audio and Visual Context Switch Indicators in a Multirobot Navigation Task. In: Proceedings of the Conference of Systems, Man, and Cybernetics, The Hague, Netherlands (To appear, 2004)
13. Wickens, C.D.: Processing resources and attention, D. Damos (ed.), Multiple Task Performance, pp. 3–34 (1991)
14. Wickens, C.D., Dixon, S., Chang, D.: Using Interference Models to Predict Performance in a Multiple-Task UAV Environment – 2 UAVs, Technical Report AHFD-03-9/MAAD-03-1, AHFD (April 2003)
15. Wickens, C.D.: The Structure of attentional resources. In: Nickerson, R.S. (ed.), Attention and performance VIII, pp. 239–257 Hillsdale, NJ

Analysis of Multilocational and Mobile Knowledge Workers' Work Spaces

Matti Vartiainen

Laboratory of Work Psychology and Leadership,
Department of Industrial Engineering and Management,
Helsinki University of Technology,
P.O. Box 5500,
FIN-02015 TKK, Finland
matti.vartiainen@tkk.fi

Abstract. The job demands and contents of knowledge workers were explored. The data was collected in three phases: first, each member filled a self-observation diary for seven days, second, each employee was interviewed individually, and, third, a workshop was organized to validate the observations and to create ideas for the development of work. The study shows that around 40 per cent of total work time is used in solitude doing tasks requiring concentration. The social network of employees is still wide, consisting of tens of people. They are contacted virtually and face-to-face. The work is pretty virtual and mobile. The study also found that the work itself is blurred. People work both in solitude, asynchronously with others, virtually online and in face-to-face collaboration with others.

Keywords: Mobile work, multilocality, knowledge work, quasi-privacy.

1 Purpose and Research Questions

This case study is aimed at deepening the understanding of mobile and distributed work. The main precondition of a distributed team's work is collaborative communication. Therefore, it is critical to find out to what degree physical, virtual and mental/social spaces support full communication and collaboration in a network of people doing project work that requires problem-defining and -solving abilities, as well as creativity. The research questions of the study are:

- For what purpose (task), how (mode of communication and collaboration) and where (place) and with whom (network) do mobile employees act and communicate?
- How do physical, virtual and mental/social spaces support, or how should they support, the team in its work?

2 Frames to Analyze and Describe Multi-locational Work

A specific system theory, the activity system approach is used as the methodological guideline for the analysis and description [9, 10, 11]. In the approach, work is studied as a system consisting of a subject using tools to process objects of work in a working

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 194–203, 2007.

context. Activity systems in working contexts are goal- and interest-driven entities, which aim at fulfilling given or self-set tasks and assignments. Work is realised through purposeful object-oriented and/or communicative actions often in collaboration with others. Subjects, as actors, are social and cultural entities such as individuals, pairs, groups, organisations, and networks. They use both concrete and mental tools to work on their objects in their respective environment, which can be characterised by its degree of complexity. The objects of work are manifested as self-set and given assignments, tasks, and goals related to them. In addition to goals, a driving force can be an interest without any exact goal, but one, which does, however, create joint actions. Because of the systemic nature of work and working, the concepts like "virtuality" and "mobility" are just aspects of activity systems.

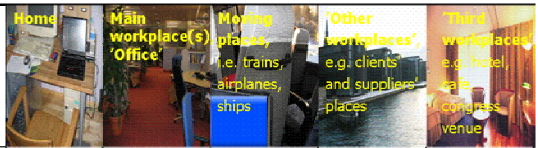
The working contexts can be outlined both from the individual and collective perspectives. From the *individual point of view*, each individual exists in a psychological field of forces that determines and limits his or her behavior. This implies and underlines the meaning of personal perceptions and interpretations of the contexts-in-use. Lewin [6] called this psychological field the "life space". It is a highly subjective "space" dealing with the world as the individual sees it. "Life space" is, however, imbedded in the objective elements of physical and social fields. As "life space" describes individual contexts, the concept of "ba" [8][1] focuses on *shared contexts*, which is useful for differentiating various spaces in collaborative work. *Ba* refers to a shared context in which knowledge is created, shared and utilized by those who interact and communicate there as often happens in knowledge work. *Ba* unifies the *physical space*, such as an office space, the *virtual space*, such as e-mail, and the *mental or social space*, such as common experiences, ideas, values, and ideals shared by people with common goals as a working context. The workplaces that virtual mobile employees use are analyzed in the following sections by using these shared space-categories (Figure 1).

Physical spaces The physical environments that employees use for working are divided into five categories: (1) home, (2) the main workplace ("Main office"), (3) moving places, such as cars, trains, planes, and ships, (4) a customer's and partner's premises or own company's other premises ("other workplaces"), and (5) hotels and cafés etc. ("third workplaces"). The use of physical places can be described by different indicators, such as their distance from each other (near – far), their number (one – many), and the frequency with which they change (seldom – often). The indicators are then used in modeling various types of distributed and mobile work units. A physical place itself can move, for example, a car, a train, or an airplane. This type of working in many places is called multilocational work [7].

A *virtual space* refers to an electronic working environment or virtual workspace or collaborative working environments. The internet and intranet provide a platform for working places for both simple communication tools, such as e-mail, and complex ones, such as collaborative working environments, which integrate different tools like e-mail, voice, videoconferencing, group calendar, chat, document management and presence awareness tools. The use of virtual workspaces can be analyzed and described by focusing on connections, devices and services and on their purposes, functionality and usability.

---

[1] 'Ba' roughly means 'place'. The concept was originally proposed by the Japanese philosopher Kitaro Nishida (Nishida, K. An Inquiry into the Good, 1921) and further developed by Shimizu (Shimizu, H. 1995, Ba-principle: new logic for real-time emergence of information, Holonics, 5, 1, 67-79.) (See Nonaka et al., 2000, 14).



Physical Spaces - Settings - Arenas - Environments - Tasks	Home	Main workplace(s) 'Office'	Moving places, i.e. trains, airplanes, ships	'Other workplaces', e.g. clients' and suppliers' places	'Third workplaces' e.g. hotel, cafe, congress venue
Virtual Spaces - Connections - Devices - Services - Purposes - Functionality	PC, phone, Internet, broadband, wlan	Intranet, communicat-ion and collaboration systems	Mobile devices	Intra- and extranet, Internet	Laptop, intranet
Mental and Social Spaces - G&O and HRM issues	Tranquility, well-being **family**	Shared goals and values, 'stress', **peers**	Change and solitude, **strangers**	Trust, **partners**	Interruptions, **mostly strangers**

Fig. 1. Types of workspaces in multi-locational and virtual work [12]

The combination of physical work settings and virtual space has been called a "*workscape*" [4]. The term "workscape" refers to the "layers of where we work", i.e. the constellation of (1) real and virtual work settings, i.e. furniture + IT, within (2) particular spaces, i.e. meeting rooms, project areas, cafés etc., that are, again, (3) located in a specific environment, i.e. office building, city district, street, home, airport, bus etc. They together form a hybrid work environment.

A *mental/social space* refers to cognitive constructs, thoughts, beliefs, ideas, and mental states that employees share. Creating and forming joint mental spaces requires communication and collaboration, such as exchanging ideas in face-to-face or virtual dialogues. Mental/social spaces are usually studied by collecting individual perceptions, attitudes and conceptions, and then analyzing their contents. Network analysis is also used to show the relationships of individual members like "liking" and "not liking" or "helping" or "not helping".

3 Methods and Data

Object of Analysis The core group (MO) of mobile employees consists of people employed to promote the mobilization of a company businesses [5, 12]. MO has its extended group, people with whom they are working for their clients (Figure 2).

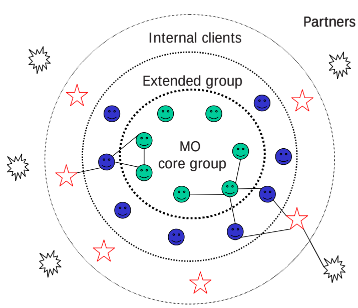


Fig. 2. The MO Team and a sketch of its network

Collection of Data. The data was collected in three phases: first, each member filled a self-observation diary for seven days, second, each employee was interviewed individually, and, third, a workshop was organized to validate the observations and to create ideas for the development of work. In addition, each participant was asked to save the messages sent and received by the different media.

Analysis of Data. First, the overall view of a working day and a week was formed by analysing the diary data at the individual and day levels. Second, the task contents and the forms of asynchronous communication while working in solitude were analyzed. The task contents were grouped following the generic knowledge task categories presented by Harrison et al. [4]. In addition, the complexity of tasks, i.e. their cognitive requirements, was described by categorizing them from "routine tasks" to "creative tasks" based on Hacker [3]. The frequencies of different forms of asynchronous work-related communication, i.e. e-mails, SMS, MMS and communication chains, e.g. one-to-one chat, were counted and the purposes of their use described. Third, the frequencies of synchronous meetings and their purposes were categorised by using the list of generic co-operative tasks [1]. Fourth, the physical spaces in use during the working week were analyzed by using five categories as described above. Time used in each place was counted, and the route during the working week was described. Fifth, the information and communication (ICT) tools used were first listed by person and then summed up to describe the whole tool set of MO members.

4 Working Spaces of a Mobile Team

In all, employees worked in physical solitude, on average, for 22 hours of the average total weekly working time of 52 hours. Working days were "blurred", meaning that when working in solitude as well as when working with others employees were often interrupted by virtual collaborative and communicative actions. Other events like official and ad hoc face-to-face meetings, lunch and coffee breaks, moving from one place to another and family affairs interrupted the day's run as well.

The workday of employees is seen as a series of work and communicative actions as episodes taking place in hybrid workspaces that are imbedded mixtures of physical, virtual and social settings. The settings are, in practice, intermingled and change dynamically as an employee during the day flexibly moves from one episode to another working some time physically alone in solitude and then with many others face-to-face. Working in solitude does not mean just "working alone in privacy", because working is affected either by self-initiated virtual outgoing contacts with others by phone and online chat, or externally by an incoming flow of requests and questions by e-mails and text messages. This is a transitional stage between deep concentration in flow and fully social polyphonic events, which is referred to here as the stage of "pseudo-privacy" [2]. It was also found that, at the other end, in large face-to-face gatherings, the mode of working returned to that of "working alone", e.g. writing documents and reading and sending e-mails while sitting in the meetings.

4.1 Forms and Contents of Asynchronous Virtual Communication

In all, the five respondents received, trashed and sent 1239 e-mails during the week in question. E-mails were considered to work best in exchanging information and opinions and worst in creating ideas, problem solving and getting to know each other. There were some doubts as to their usefulness for persuasion, bargaining and resolving disagreements. Disagreements were resolved to some extent by e-mails, however, it was remarked that resolving disagreements was easier to do face-to-face. The problem of e-mails is their number.

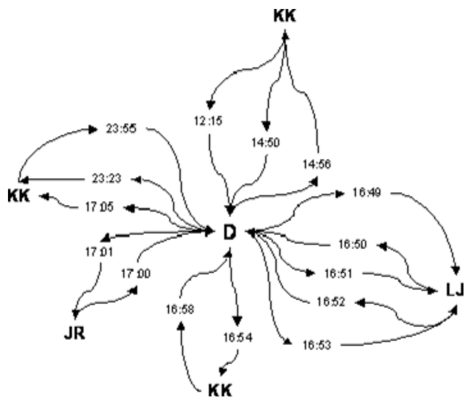


Fig. 3. An example of simple ego-centric SMS message chains of the respondent D during one day (KK = mr. Karl Kirman, LJ = mr. Larry Jaatinen, and JR = mr. Jack Richman)

In all, 121 SMSs were received and sent during the week. SMS was used for asking and answering detailed questions, exchanging information, e.g. “Where are you”, “Which room we meet in?”, informing, e.g. “The call is postponed due to meetings”, reminding; confirming; suggesting, checking flights, getting information, e.g. “Weather forecast in Singapore”. It is noteworthy that the use of multimedia messages (MMS) was rare among the respondents; only two instances when MMS were used were reported. This raises the question about their role and benefit in work.

Exchanging e-mails often develops as a communication chain (Figure 3). All kinds of communication chains are used: from very short-cycle chains where several messages are exchanged in a couple of minutes, chains lasting couple of hours, and communication chains lasting weeks and even longer. The throughput time of a matter in a chain may be even months. The mixture of media used in a chain may also vary from one media to multimedia use. For short-cycle communication chains, it is usual to use the same media, for example SMS.

4.2 Forms and Contents of Synchronous Virtual Communication

The data shows that, during the observed week, the most frequently used media for real-time communication were individual calls and small official and ad hoc face-to-face meetings. Most respondents also used chat and call conferences on daily bases.

Most employees also had large teleconferences during the week. No one reported having videoconferences. The usefulness of teleconferences seems not to be good for creating new ideas, problem solving and resolving disagreements. It was mentioned that teleconferences work well when all participants participate virtually, but that problems occur when only part of the participants take part via telephone.

Online chatting was quite popular. It was used mostly for quick checking. It was said to suit arguing and small problem solving very well. It was also used for exchanging information and opinions, sometimes also for generating ideas. Some had also used it to get to know somebody, bargaining and persuading clients, and even resolving disagreements. Problem solving was accomplished a lot during chat: someone asks something and then people start discussing the matter. Often along with problem solving, people also generated new ideas.

Small as well as large teleconferences were mostly used for exchanging information and opinions. Often teleconferences were supported by online chat and documents were shared and handled. The use of persuasion, bargaining, problem solving and generating ideas was rated low. In brainstorming you need to move and draw, it was said. Teleconferences do not support these requirements.

4.3 Forms and Contents of Synchronous Face-to-Face Communication

Face-to-face meetings with teleparticipants were organized fairly often. Quite unanimously, interviewees remarked that these meetings with with a number of others participating through teleconference did not work well.

Small face-to-face meetings were very popular and successful. They were especially used for exchanging opinions and information and for persuasion. In fact, they were used for all of the purposes categorized earlier. It was said that one-to-one meetings were for deep discussions, problem solving and creating ideas.

Larger (2-4 people) meetings were used mostly for collaboration. They seem to be best for problem solving and generating ideas. They were also thought to work well when exchanging opinions and information. Eventually they could also be used for bargaining. Small official face-to-face meetings were also sometimes used for generating ideas (meetings aimed at planning), sketching and writing.

Large meetings (5-10 participants) were used for exchanging information, though it was often seen to happen one-sidedly. They were also for lobbying (persuasion) and sometimes getting to know others.

Really large meetings (more than 10) were mainly used for distributing information and exchanging opinions, sometimes for self-presentations of new employees. Another type is a large meeting, where something is shared, and opinion or advice is asked. In large meetings, persuasion and lobbying were also seen to occur. A kind of paradoxical thing seems to happen in large meetings: people start to work "in solitude" instead of collaborating together. In large meetings, people start to concentrated on own tasks and work asynchronously: reading and sending e-mails and SMS, chatting, reading documents, and writing them.

One-to-one ad hoc meetings were used for exchanging information and opinions and for problem solving. They usually started from exchanging opinions and information. They were often arranged to check the status of some project. They were said to easily change into problem solving and drift into decision-making. In official

meetings, specific tasks were dealt with, but in ad hoc meetings things were taken to another level (“meta level”). Larger ad hoc meetings (3-5 participants) were said to be more problematic. They were more difficult to arrange, as people were so busy and distributed in different places.

4.4 Multi-locational Work and Moving Around

The employees worked in company’s premises 55 per cent of their total working time. After the main office, home was the most used place. The use of home for work varied from eight per cent to 61 per cent.

The work of the interviewees was multi-locational and mobile on several levels: interviewees traveled once in a while, for example, to Boston, Dallas and Singapore. However, most traveling took place in Finland, especially around company’s premises in the metropolitan area. There was also micro-mobility in the main office’s meetings rooms and corridors. In addition, airports, shops and even a doctor’s waiting room were mentioned as working places.

Interviewees moved a lot inside the main office. “Office” was considered to be the whole main office complex, the campus. It was the place to meet other people. Most of the meetings took place there, but also the company’s main building was used a lot. The study revealed that half of the time was used in meeting rooms. Auditoriums were used as places where you can sit either as one of the audience or you may participate more actively. Calls were made in flexi-spaces and corridors. Interviewees mentioned that speaking and calling in flexi-spaces were problematic because of causing interruptions to others’ work. Café or tea machines, corridors, rest rooms, sofa corners and elevator fronts were used a lot.

There were not many comments on working in moving places. Sometimes calls were said to be made in cars and taxis. Phone calls in public transportation were rare because of business security reasons. Short trips in busses were utilized to read and send e-mails. Airplanes and airports were used too, while coming back from long journeys. Interviewees mentioned that during a flight it was a relief to do some work, on documents, for example, or answering mails to send them after landing.

There were several references to home as a workplace. Home was thought to be the place where one could concentrate on demanding tasks. The possibility of working at home was considered to be an advantage, but it was also seen to cause difficulties in separating work and family life.

“Third places” like cafés in cities, shops, and university premises were mentioned and evaluated only seldom. It seems that they are not in very general use. Some also mentioned nature as a place for thinking.

4.5 ICT Tools in Use

The virtual environment in use was versatile. Table 1 shows the interviewees’ description of their ICT tools, their usefulness and ideas as to how to improve the use. Basically, the tool set for everyone was standard. Chat, laptop and e-mail were mentioned as the most beneficial. On the other hand, it was mentioned that improving employees’ competencies could develop their use.

Table 1. Media in use, indispensable media and ideas to improve media use

Media in use	Indispensability	Ideas to improve
- Basically the same standard tools. - Phone (calls, sms, smash, push, e-mail, chat, intranet, internet, contacts, calendar), laptop (e-mails, power point, sametime chat, teleconference, online presence, Trackview, Lotus Notes-based Teamroom for documents, intra- and internet, mindmap, hour reporting, MsOffice), paper documents, PostIt, Locus magazine, Irma for showing slides with phone. - Most people have a home office connection.	- It depends on what you are doing. For making things happen, chat is the number one. Although other person is in a meeting, by using chat you can manage things. Chat is unnoticeable. Nobody waits for long answers, just quick ones. Instead of putting things into mailbox, it is good. - If PC is taken off, everything becomes paralyzed. - E-mail is the best and worst media: quick, multifaceted, mobile access but used too much.	- E-mails are used too much. Instead all the material could be put to Teamroom and the e-mail would just include links. - There are too many meetings. Because e-mails are used so much, other tools are suffering. - Needs for videoconference if properly arranged. It is a hindrance to get opposite side to right place. - Centralized calendar is a problem, because you cannot know where the others are, and traveling may take some time. - Virtual meetings tools should be more flexible. - Communication tools are not problems as such; people just do not know how to communicate.

5 Discussion

The job contents of the employees are demanding both cognitively and socially. Around 50 per cent of the work includes thinking and creativity demands. So, team members can be genuinely called knowledge workers. Around 40 per cent of total work time is used in solitude doing tasks requiring concentration. The social network of employees is wide, consisting of tens of people. They are contacted virtually and face-to-face. During seven days, each person received and sent 24 SMSs, had 26 small and large face-to-face meetings, worked in solitude 35 times, received and dialled 29 calls, and received, sent and trashed 248 e-mails on the average.

The study revealed that MO members' work is pretty virtual and mobile. The work MO members do is not tied to any one location, nor is it tied to any specific time. Work is done at the main office, company's other premises, home, sometimes in public spaces, in cars, and, for instance, in the offices of clients and colleagues. Work starts early in the morning, and often the days end with some work-related tasks late in the evenings. Many work on Sundays too.

The study also found that the work itself is blurred. People work both in solitude, asynchronously with others, virtually online and in face-to-face collaboration with others (Figure 4). All are essential for success of work, but contrary to the traditional thought that these modes of work can be separated. It is rather difficult to separate working in solitude from collaborative work. Working in solitude might mean virtual asynchronous collaboration with others and presence in collaborative meetings may be used to write some documents in solitude. Thus the nature of work seems to have become all the more blurred at several levels.

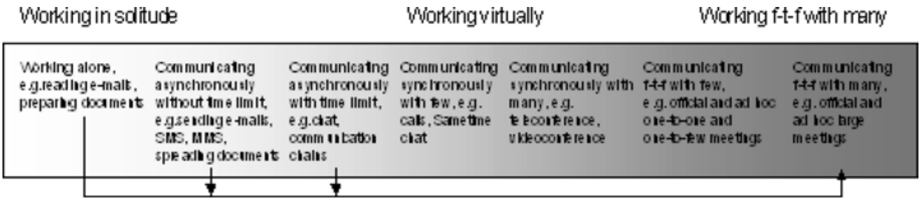


Fig. 4. A workday is a mixture of working spaces

The employees worked on the company's premises 55 per cent of their total working time. Only one of the employees worked more than 50 per cent of the time at the main workplace, and another one more than 50 per cent at home. Home was the most used place. The use of home for work varied from eight per cent to 61 per cent.

Employees are multi-locational and mobile at several levels. Some of them were mobile at the global levels, very often they were Campus mobile, visiting different sites of the extended enterprise, and they were micro-mobile in the main office's meetings rooms and corridors. Travelling itself, especially between company's offices, was considered a nuisance: necessary but uncomfortable and unproductive.

One of the aims of this study was to explore how spaces, ICT and human resources manage to support the contemporary work of employees. It was found that, at the moment, people doing virtual and mobile work experience challenges that these support functions could – and should – address. Open space offices did not allow people to concentrate properly, thus home had become an important place for work that demanded concentration. The spaces that enable both concentration on individual tasks as well as on collaboration are needed. It was also mentioned that virtual collaborative working environments could be further developed to better support real, virtual and mobile work. At the moment, proper videoconference facilities are not available and there are some problems having successful teleconferences, especially if some participated virtually and others face-to-face. Also, the new blurriness of work, diffusion of work into different places, different times and different networks were considered troubling at times. There was no clear difference between working time and leisure, and thus it was thought to be difficult to really relax and forget work-related issues. As work seemed to dominate the week of mobile employees, some clear guidelines for alleviating stress and managing new kinds of work were needed.

References

1. Andriessen, J.H.E.: Working with groupware. In: Understanding and evaluating collaboration technology, Springer, London (2003)
2. Becker, F., Sims, W.: Managing uncertainty. Integrated portfolio strategies for dynamic organizations. Cornell University, International Workplace Studies Program. Ithaca, NY (2000)
3. Hacker, W.: Allgemeine Arbeitspsychologie. Psychische Regulation von Wissen-, Denk- und körperlicher Arbeit. Hans Huber, Bern (2005)
4. Harrison, A., Wheeler, P., Whitehead, C. (eds.): The Distributed Workplace. Spon Press, London New York (2004)

5. Lattanzi, M., Korhonen, A., Gopalakrishnan, V.: Work goes mobile. Nokia's lessons from the leading edge. John Wiley & Sons, England (2006)
6. Lewin, K.: Need, force and valence in psychological fields. In: Hollander, E.P., Hunt, R.G. (eds.) Classic contributions to social psychology, Oxford University Press, London (1972)
7. Lilischkis, S.: More yo-yos, pendulums and nomads: trends of mobile and multi-location work in the information society (STAR (Socio-economic trends assessment for the digital revolution) Issue report no 36 (2003), URL: http://www.databank.it/star
8. Nonaka, I., Toyama, R., Konno, N.: SECI, ba and leadership: a unified model of dynamic knowledge creation. Long. Range Planning 2, 5–34 (2000)
9. Vartiainen, M.: Mobile virtual work – concepts, outcomes and challenges. In: Andriessen, J.H.E., Vartiainen, M. (eds.) Mobile Virtual Work: A New Paradigm?, pp. 13–44. Springer, Heidelberg (2006)
10. Vartiainen, M., Andriessen, J.H.E.: Mobile virtual work: what have we learnt? In: Andriessen, J.H.E., Vartiainen, M. (eds.) Mobile Virtual Work: A New Paradigm?, pp. 369–386. Springer, Heidelberg (2006)
11. Vartiainen, M., Hakonen, M., Kokko, N.: Emergence of mobile virtual work: concepts, outcomes and challenges. In: Sachse, P., Weber, W.G. (hrsg.): Zur Psychologie der Tätigkeit. Huber, Bern, pp. 298–327 (2006)
12. Vartiainen, M., Hakonen, M., Koivisto, S., Mannonen, P., Nieminen, M.P., Ruohomäki, V., Vartola, A.: Distributed and mobile work – places, people and technology. Gaudeamus, Helsinki (2007)

Are Computers Capable of Understanding Our Emotional States?

Min Cheol Whang[1], Joa Sang Lim[1], Kang Ryoung Park[1], Youngjoo Cho[1], and Wolfram Boucsein[2]

[1] Sangmyung University, 7 Hongji-dong, Jongro-gu, Seoul, Korea
[2] University of Wuppertal, Max-Horkheimer-Str. 20, 42119 Wuppertal, Germany
{whang, ycho, jslim, parkgr, ycho}@smu.ac.kr,
boucsein@uni-wuppertal.de

Abstract. As emotion provides an important clue for communication, the computer needs to be more sympathetic to users' commands in the context of their emotion. A challenging attempt has been made to develop an emotional computer, which reads such physiological signals as photoplethysmogram, electrodermal activity and skin temperature and analyzes them online with a rule base into human emotion. We adopted a two-dimensional emotion model and a number of empirical studies have been conducted to find out some valid physiological parameters and to map them with nine categories of human emotions. Some research challenges were found that need to be addressed until the emotional computer comes to the market.

Keywords: Emotion in HCI, Emotional computer, Psychophysiology, Emotional mouse.

1 Introduction

Research into human computer interaction (HCI) has concentrated in the past on cognitive issues such as effectiveness and efficiency, easy-to-understand menus, and appropriate temporal structures [1], [2]. While interacting with a computer, however, users experience emotion, which certainly affects HCI. Emotion has been often regarded as an important context from a communication point of view [3]. However, it's role has been less clarified in HCI [4]. Positive emotions often result in faster, simpler and more flexible information processing strategies, while negative ones may result in more systematic strategies but also slow down HCI substantially [5], [6]. Taking emotion into account as a key for a satisfying HCI, a number of studies addressed this problem recently [7], [8]. To recognize human emotion automatically, HCI studies rely on diverse measures such as facial expressions, spoken words, body gestures or physiological signals [9], [10]. These measures may be subject to various limitations such as environmental conditions (e.g. lighting, audio recording) or obtrusiveness. The present study aimed to continuously record the user's emotion during HCI by means of psychophysiological methods as unobtrusively as possible, perform an online analysis of the user's emotional state and respond appropriately to it.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 204–211, 2007.

2 Methods

2.1 Emotional Model

This study has adopted a two-dimensional model of emotion suggested by Larsen and Diener [11]. All emotions can be located on circumplex in the two-dimensional space made up from valence and arousal. Emotions are classified in each of the four quadrants made up by the two bipolar axes: pleasantness (P) vs. unpleasantness (U) and arousal (A) vs. relaxation (R). The four quadrants were named after the combinations of the axes (U-A, P-A, U-R and P-R) as shown in Fig. 1. In an earlier study, the present authors successfully probed the usability of the two-dimensional system for monitoring the basic emotional states by means of psychophysiological measures [12]. The four basic emotions were induced by different combinations of olfactory and auditory stimuli. Over the emotional treatment, recorded were electrocardiogram (ECG), electrodermal activity (EDA) and four channels of electroencephalographic activity (EEG) along with regularly measured subjective emotional states. Multivariate combinations of psychophysiological parameters figured out to be superior to any single measure in detecting the subject's emotional state.

2.2 Physiological Measures of Emotion

Any physiological approach to HCI would bring about obtrusiveness, which is the most evident for EEG. Thus, autonomic nervous system measures were selected for these needs. A measurement device called "emotional mouse" (see Fig. 2) was specially constructed for recording three measures that cover the most salient aspects of autonomic activity, i.e., electrodermal activity (EDA) as an indicator for pure sympathetic activity [13], skin temperature (SKT) for parasympathetic activity and photoplethysmogram (PPG) for arousal and orienting. EDA is a slowly changing signal and rising of its level means an increase of tension. SKT is also a slow signal, but rising means an increase in relaxation. PPG is a relatively fast signal with a peak about every 0.8 second, which is correlated with the heart beat. Thus, heart rate (HR) can be derived from PPG as an indicator for arousal vs. relaxation, and the modulation of the PPG volume amplitude provides a measure of orienting.

2.3 General Outline of the Algorithm

Built into the computer was an emotion processor to measure and analyze physiological signals continuously recorded from the users. Sensors, which were designed to be wearable and comfortable with a minimal burden from their attachment, were used for recording. The signals were amplified, filtered, and digitized by a data acquisition system designed as small and light as possible. Automatic noise detection and removal should be performed according to pre-defined specific algorithms. The pre-treated signals were processed to extract appropriate physiological parameters that were qualitatively normalized based on a defined state of emotional neutrality. These normalized data were referred to a special algorithm or rule base for evaluating the user's current emotional state. Both measured and

processed data were stored in the database, which can potentially upgrade the algorithm or rule base. Since the subjective report played a central role in determining the qualitative aspects of any emotion, it was necessary to confirm the emotion evaluated by recording the user's subjective emotional state during a data acquisition phase and save it together with the physiological measures in the database. These confirmed data were used to upgrade the database for the algorithm and/or the rule base. Furthermore, the system was trained for an evaluation of an individual user during a learning process. An emotion indicator was used to present the user's emotion on the screen. In case of a negative emotion being detected, the program initiated a service of helping the user to have a more positive emotion.

2.4 Neutral Band for Reference

Computer users stay comfortable without any noticeable changes in their emotion with respect to arousal and valence during most of the time. Thus, the four quadrants defined by two-dimensional model as outlined in 2.1 are not sufficient for explaining a region for the neutral state of emotion. However, such a region of neutral emotion is needed as a reference state, from which the non-neutral emotions should be defined. Therefore, the emotional model was refined by adding a neutral band to the original emotions as shown in the left panel of Fig.1.

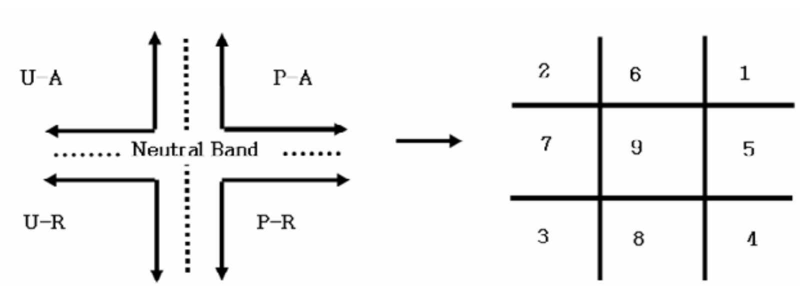


Fig. 1. Neutral Band of Emotion

Taking such a neutral band into account for the boundary among each pair of the quadrants and having an additional neutral-neutral field in the center increases the number of emotions from 4 to 9, as can be seen in the right panel of Fig. 1.

The neutral band was defined separately for each physiological measure. Because the sensitivity of the measures may vary over different persons, neutral bands were defined separately for each individual. A relatively narrow neutral band corresponds to high sensitivity but low accuracy of determining the emotion, and vice versa. During the system's adaptation process, subjective assessments of emotion were used to automatically adjust the neutral bands wider or narrower. Therefore, the neutral band was different in the physiological signals within a user and among users at the same time.

2.5 Rule Base for Emotion Determination

To minimize any individual differences in physiological readings, a neutral band was employed as a reference as discussed earlier. Given this, the percent variation of incoming physiological signals from the defined neural band was calculated as shown in Equation 1.

$$E = (C-N)/N \tag{1}$$

E is expressed in percent and equals the increase or decrease of each signal from the neutral state (N) to the current state (C). These normalized values of the physiological signals were used to set the rule base which used to determine emotions. The rule base used three different entries for signal changes: an increase (+E) or decrease (-E) from the neutral band, or a zero value (0) in case of no deviation from the neutral band. Given these three entries (+E, -E, and 0) and three physiological measures (PPG, EDA, and SKT) resulted in 27 different patterns. Six combinations were eliminated because they showed an opposite direction for physiological reasons. The all-zero case was considered as a reference and therefore also eliminated, so that 20 different combinations of physiological changes remained for setting up the rule base. As a consequence, different patterns may result in the same emotion.

Besides individual differences in the neutral band, there are also individual differences to be taken into account for the rule base. For example, a 5% increase of SKT, a 10% decrease of EDA and a 15% increase of PPT from the neutral band may indicate relaxation in one person, while 20% increase of SKT, a 5% decrease of EDA and a 10% increase of PPT may indicate relaxation in another person. Therefore, the values of the expression E were individualized for all nine emotions. The personal set of rule base was automatically updated by using the subjectively reported emotion.

2.6 Subjective Confirmation of Emotion

The physiological signals do not fully account for the variation in the emotional states of individual users and thus, subjective confirmation of the emotion is required. In addition to the psychophysiological measures of emotion, subjective ratings of perceived arousal and valence were obtained in regular intervals. A pop-up window was presented, asking the subject to click on one of three buttons, representing relaxation, neutral arousal and high arousal state. Thereafter, another window asked for unpleasant, neutral or pleasant feelings. To form a stable data base for each individual, a subjective evaluation was performed every two minutes during the initial stage of emotion evaluation. If the incoming physiological signals were not in accordance with the subjectively reported emotion, the rule base was adjusted. After having reached a stable data base, the distances between subjective evaluations increased to five minutes or more.

2.7 Hardware and Software Development

To obtain the least noisy physiological readings of PPG, EDA and SKT, an emotional mouse was carefully designed to establish a firm contact between the bio-sensors and human palm. Ten design factors were taken into consideration during product development. The width (9-10 cm) and the length (18-19cm) of the emotional mouse were determined in reference to Korean anthropometric data of twenty year-aged in 1992.

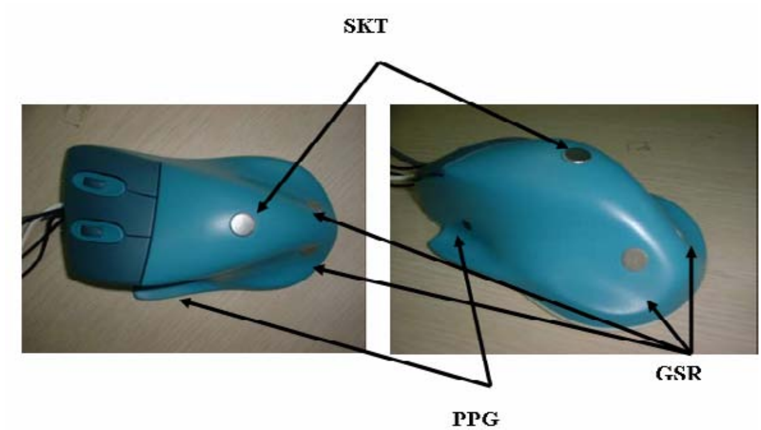


Fig. 2. Emotional Mouse

Fig. 2 shows the sensor location in the emotional mouse. It uses three curvatures that reflect the natural form of the hand; the thenar-hypothenar curvature for EDA (Fig. 27, [13]), the curvature of the inner palm for SKT, and a special thumb holder for the PPG, preventing the thumb from moving up- and downward.

A customized integrated circuit board connecting the emotional mouse and the PC was constructed for acquiring, filtering, amplifying, and AD converting PPG, EDA and SKT signals. It communicated with the PC via RS 232. Since data recording and processing lead to an excessive consumption of CPU, causing the computer to operate slowly, client-server architecture was employed. The client PC had only the functions of recording physiological signals and indicating the results of the emotion evaluation. The server performed the emotion analysis and its evaluation, and also the updating of both the rule base and the data base. The information of the subjective emotional state was sent to the server together with the physiological signals. The server set and updated the neutral emotion band and the rule base, using new qualitative physiological data. The actual emotion was determined and sent to the PC for display. In addition, the PC indicated the results of emotion evaluation by a pop-up emotion indicator, consisting of bars graphs, verbal explanations and a schematic face icon (Fig. 3).

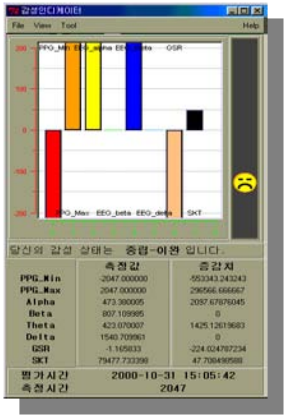


Fig. 3. Pop-up emotion indicator

For the icon, the mouth angles can be down, neutral or up, combined with three different orientations of the eyebrows, resulting in nine different icons that represent the different emotions. The user can decide to see additional detailed information, given in numbers that represent results of physiological measures. The icon stays on the screen continuously until being updated.

The subjective confirmation was used to identify the current emotion. It was correlated with the incoming physiological data and also continuously used to re-evaluate the neutral emotion. During on-line processing, the data base was continuously developed and updated. It consisted of the stored physiological parameters together with the E values for the 9 emotions. This data accumulation provides a continuously improving emotion evaluation that also takes personal characteristics into consideration.

2.8 Negative Emotion Recovery Service

The negative emotion may certainly be the state where the users do not want to stay on. Thus, the emotion computer should be capable of providing a service to mitigate negative emotion, once it is detected. This service was elicited by detecting the unpleasant emotions 2 and 3 (see Fig. 1). First, a pop-up window was presented, asking the subject to confirm the acceptance of this service. If accepted, the subject’s favorite multimedia data base was provided, offering his/her favorite music, video clips and the possibility to get rid of the negative emotion via playing a short video game. For this, the multimedia preference of the user was determined prior to using the emotional computer system.

2.9 Empirical System Verification

The emotional computer has been empirically verified as to the consistency between qualitative emotion as assessed by the emotional computer and subjective reports of emotion. Five voluntary subjects took part in the experiment. The subjects rated their emotions 100 times for three days. Results showed 70-90% identification of the

subjective emotion by the psychophysiological evaluation. The results will be presented in detail elsewhere. In general, the assessment was more accurate for arousal than for valence.

3 Discussion

During the past, research has focused mainly on technical aspects of user interfaces such as the customization of graphical interfaces and the design and behavior of interface objects [1]. More recently, there has been a growing interest in the principle of user-centered design in user interfaces, including not only users' cognition but also their emotion into computing environment [4] [14]. Thus, the time has come to challenge the behavioral issues of emotion in the context of HCI. Most pertinent questions will be: "what influences does human emotion have on the performance of computer tasks?" and "what aids should be provided to let users stay with their comfortable feelings?"

The first aim of the present study was to propose a theoretical framework for an emotional computer. It consisted of three functional components of (1) physiological measurement, (2) emotion recognition and (3) adaptive feedback. Based on these principles, we successfully developed an emotional computer with the aid of a specially-designed mouse for sensing the user's autonomic responses. As physiological signals, PPG, EDA, and SKT were recorded and forwarded to a server for data processing. The bio-signals collected were then analyzed to recognize the user's emotional state. The evaluation of emotion was individualized according to updating the neutral band and the rule base by means of subjective confirmation. The emotional computer was designed to provide a multimedia service in case of negative emotion to make it positive. An empirical verification with five student subjects showed 70-90% accuracy in identifying the subjective emotional state. Given the amount of data collected, these results are encouraging compared to the accuracy of the earlier study [12].

In relation to the theoretical framework provided, a number of research issues can be raised. Firstly, the adaptive nature of emotional computers assumes their actions in good tune with the subjective emotion of users. Such adaptability is to a great extent trusted to the signal processors of emotional computers. Thus, research should be made to determine under what conditions mathematical models perform robustly (e.g., neural networks [9]). Secondly, the apparatus to be attached to the human body should be kept to a minimum size and be unobtrusive to users. The device may include some sensors and transmission systems required to measure and transfer the physiological data, which can be quite obtrusive to users [10]. Lastly, user interfaces should also be designed to provide appropriate action to alleviate the uncomfortable feeling of a user that may occur in the course of HCI [8]. It should be noted that too much care of unstable emotion may lead to irritation. Thus, in addition to constant monitoring of physiological emotion, the subjective and behavioral state of the user's emotion should be registered to keep track of the appropriateness of the adaptive behavior of user interfaces. The distinctive feature of an emotional computer is its adaptiveness of user interfaces, which plays a role in differentiating it from ordinary computers. The development of adaptive user interfaces is far from simple and

requires interdisciplinary efforts. First of all, we need to know what provokes uneasy feelings of users and their effect of emotion on computer work [15]. The emotional computer will be continuously developed for next version which is wireless, mobile and portable, for future applications in ubiquitous computing environments.

References

1. Shneiderman, B.: Designing the User Interface: Strategies for Effective Human-Computer Interaction, 3rd edn. Addison Wesley, Menlo Park, CA (1998)
2. Boucsein, W.: The use of psychophysiology for evaluating stress-strain processes in human-computer interaction. In: Boucsein, W., Backs, R.W. (eds.) Engineering Psychophysiology: Issues and Applications, pp. 289–309. Lawrence Erlbaum Associates, Mahwah, NJ (2000)
3. Oatley, K.: Emotion. The Psychologist, 285–288 (1998)
4. Hudlicka, E.: To feel or not to feel: The role of affect in human-computer interaction. International Journal of Human-Computer Studies 59(1-2), 1–32 (2003)
5. Norman, D.: Emotional Design: Why We Love (Or Hate) Everyday Things. Basic Books, New York (2004)
6. Forgas, J.P.: Network theories and beyond. In: Dalgleish, T., Power, M. (eds.) Handbook of Cognition and Emotion, pp. 591–612. John Wiley & Sons, Chichester (1999)
7. Picard, R.W., Klein, J.: Computers that recognise and respond to user emotion: theoretical and practical implications. Interacting with Computers 14(2), 141–169 (2002)
8. Partala, T., Surakka, V.: The effects of affective interventions in human-computer interaction. Interacting with Computers 16(2), 295–309 (2004)
9. Fragopanagos, N., Taylor, J.G.: Emotion recognition in human-computer interaction. Neural Networks 18(4), 389–405 (2005)
10. Haag, A., Goronzy, S., Schaich, P., Williams, J.: Emotion recognition using bio-sensors: First steps towards an automatic system, in Affective Dialogue Systems. In: Proceedings. 2004, pp. 36–48. pringer-Verlag, Berlin (2004)
11. Larsen, R.J., Diener, E.: Promises and problems with the circumplex model of emotion. In: Clark, M. (ed.) Review of personality and social psychology, vol. 13, pp. 25–59. Sage, Newbury Park, CA (1992)
12. Whang, M.C., Lim, J.S., Boucsein, W.: Preparing computers for affective communication: A psychophysiological concept and preliminary results. Human Factors 45(4), 623–634 (2003)
13. Boucsein, W.: Electrodermal Activity. Plenum Press, New York (1992)
14. Picard, R.W.: Toward computers that recognize and respond to user emotion. Ibm Systems Journal 39(3-4), 705–719 (2000)
15. Scheirer, J., Fernandez, R., Klein, J., Picard, R.W.: Frustrating the user on purpose: a step toward building an affective computer. Interacting with Computers 14(2), 93–118 (2002)

A Review of Current Human Reliability Assessment Methods Utilized in High Hazard Human-System Interface Design

Chih-Wei Yang[1], Chiuhsiang Joe Lin[1,*], Yung-Tsan Jou[1], and Tzu-Chung Yenn[2]

[1] Department of Industrial Engineering, Chung-Yuan Christian University, 200, Chung Pei Rd., Chung- Li, Taiwan 32023, R.O.C
[2] Institute of Nuclear Energy Research, Lungtan, Taiwan
hsiang@cycu.edu.tw

Abstract. This paper has three objectives. The first is to discuss the role that human reliability assessment implemented in human systems interface design. The second is to present significant characteristics of available HRA techniques. The third is to provide our viewpoints of the applicability of HRA methods in HIS design. Generally, HRA approaches advocate seven stages, problem definition, performances shaping factors analysis, task analysis, human error analysis, effect analysis, error reduction strategies, and evaluation of recommendations. The most difficult technique is the human error probability estimation and prediction. There are four shortcomings of past human error probability estimation and prediction methods. First, the reliable data is deficient. Second, there are insufficient criteria for choosing PSFs. Third, there is a limited capacity for evaluating cognitive behavior. Finally, possible causes are ignored. To look into the above-described problems, this study reviewed present HRA methods and proposed several aspects for future HRA method development.

Keywords: Human error, Human reliability assessment, Human-system interface, Probabilistic safety analyses.

1 Introduction

System designers may find it more difficult to develop an excellent human-system interface (HSI) then ever before. One of the difficulties in developing HSI is that one must anticipate the response of users from a large space of possible system states and design options [7]. Hence, the computer based HSI design is more complicated then before. Today, the importance of HSI to reliable human performance and high hazard systems is widely recognized. In order to make sure the system is safe and effective; there are a larger number of researches discussing the topic of human performance in new HSI design. Human reliability assessment (HRA) is one of the most famous and efficient methods to assess human performance. The impact of advanced HSI by HRA on human performance is significant from these researches, as described in the following.

* Corresponding author.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 212–221, 2007.

Mosleh and Chang [13] considered HRA a scientific discipline involving the systematic application of information regarding human characteristics and behavior to enhance the performance of HSI. HRA was first introduced in the 1960s by Munger et al. [14] who developed the American Institute of Research (AIR) Data Store. Swain [20] developed the technique for human error rate prediction (THERP). THERP is now applied to generate input for probabilistic safety analyses (PSA) utilized by the nuclear power industry and National Aeronautics and Space Administration (NASA). Mosley and Chang (2004) proposed that HRA methods should include the following functions: (a) identify human errors; (b) estimate human error probabilities; (c) identify causes of human errors; (4) include a systematic process for generating replicated qualitative and quantitative results; (d) have a causal model of human errors with roots in cognitive and behavioral sciences; (e) be detailed to support data collection, experimental validation, and various applications of probabilistic safety assessment (PSA).

Human actions form a large class of error sources based on the identification of errors [3]. Kirwan [10] [11] introduced 38 human error identification (HEI) approaches of error identification, classifying them into many types of error identification approach. In general, HEI approaches advocate mostly the seven stages, that is, the problem definition, performance shaping factors (PSFs) analysis, task analysis, human error analysis, effect analysis, error reduction strategies, and evaluation of recommendations [19]. All these stages can be roughly divided into qualitative and quantitative techniques. The most difficult quantitative technique is human error probability estimation and prediction. These methods will be explained in the next section. The objective of this paper is three-fold: (1) to discuss the role HRA implemented in HSI design review; (2) to characterize the significant properties of available HRA techniques; (3) to provide our viewpoints in the applicability of HRA methods.

2 Human Reliability Assessment Techniques

This section summarizes five major HRA techniques. There are three key parts discussed in each method, including overview of this method, analytic process, and implementation on safe HSI design.

2.1 Technique of Human Error Rate Prediction, THERP

THERP is one of the most well-known HRA techniques, primarily for its quantification approach, which was proposed by Swain and Guttmann [21]. The possible external error modes (EEMs) were conceived at each step. EEMs describe what error occurred, in terms of the external and observable manifestation of the error. EEMs do not imply anything about the cognitive origins of the error, such as the controller's intention (Issac et al., 2002).

Kirwan [9] stated that THERP was operated by the following process. (a) Division of tasks into basic elements. (b) Allocation of nominal human error probabilities (HEPs) to each element. (c) Identification of the effects of performance shaping factors (PSFs) on each element. (d) Calculation of dependent effects between tasks.

(e) Modeling in a HRA Event Tree. (f) Quantification of total task HEP. THERP put emphasis on the quantification results. HRA assessors can describe human actions based on human error event tree. The task elements relationships will be clearly defined in the event tree. Then, human error probabilities can be calculated by a systematic process. Therefore, THERP can be applied to the task performance prediction while designing the HSI interface.

2.2 Human Error Assessment and Reduction Technique, HEART

HEART was developed by Williams [23]. Kirwan [9] proposed that HEART is one of the most well-known techniques currently used in UK. This technique has been verified to have practical validity in at least one major validation exercise. HEART was proven to have the ability of providing accurate numerical estimates of failure likelihood based on the practical and quantitative validity by Kirwan [9].

Kirwan [9] stated that HEART was operated by the following process. (a) Categorize tasks into one of the generic categories. (b) Allocate a nominal HEP to the task. (c) Verify the Error Producing Conditions (EPCs: effectively PSF) which will influent task reliability. (d) Decide the Assessed Proportion of Affect (APOA) for each EPC. (e) Compute the task HEP. This technique is more flexible and rapid then THERP. On the other hand, the data gathered by THERP is more reliable then HEART. THERP can be utilized whenever there is a main risk sensitivity HEP needed estimating. Then, HRA assessor can use HEART as a 'screening' technique to verify the analytic results of THERP.

2.3 Cognitive Reliability and Error Analysis Method, CREAM

The cognitive reliability and error analysis method was proposed by Hollnagel [5]. CREAM put emphasis on defining and analyzing the causes of human errors. The theoretical background of CREAM is the classifications of error modes and elements of humans, technology, and organization (MTO). MTO was combined with person-related factors, system or technology-related factors, and organization-related factors.

Kim [7] stated that CREAM was operated by the following process. (a) The target task is selected. (b) The task is analyzed by a hierarchical task analysis (HTA). (c) Common performance conditions (CPCs) are assessed. (d) The cognitive demand report is established to identify cognitive functions. (e) The possible control mode is recognized. (f) The possible cognitive function failure is identified. (g) The cognitive failure probabilities for each task element and complete task can be calculated. CREAM can be used to validate and verify the HSI design and improvement. There are three major advantages of using CREAM, including (a) maximizing the human performance capabilities; (b) minimizing the human error probabilities; and (c) maximizing the recovery possibilities of human errors.

2.4 Human Error Rate Assessment and Optimizing System, HEROS

The human error rate assessment and optimizing system (HEROS) was proposed by Hollnagel [5]. Fuzzy set theory (FST) was used for representing data uncertainties and describing the general human behavior. The knowledge base of HEROS is derived from the results of generally valid ergonomic and psychological studies. These results

are expressed by fuzzy variables. Therefore, it is not necessary to use existing databases, e.g. human error database of THERP.

The following descriptive processes are concluded by Richei et al. [16]. The human error probabilities can be calculated by these processes. (a) Definition of the present environment condition. (b) Definition of the target action. (c) Review of related documentation. (d) Definition of active personnel. (e) Evaluation of management. (f) Creation of the action sequence, that is, task analysis. (g) Representation by a fault tree. (h) Evaluation of PSFs. To identify the influences of PSFs on task elements can verify the influences on actions. (j) Calculation of human error probabilities. HEROS provides a process which can be applied for the qualitative and quantitative assessment of human errors in HSI. The analytic process considers the related ergonomic and psychological studies. These studies focus on management, human-machine interface, and working environment. Besides, the most noted characteristic of HEROS is to implement a rule-based expert system by applying the fuzzy set theory (FST).

2.5 Human Error Risk Management for Engineering Systems, HERMES

The human error risk management for engineering systems (HERMES) is proposed by Cacciabue [2]. The purposes of HERMES include: first, a ‘roadmap’ for selecting and applying the most appropriate human factors approaches for specific problem. Second, the ‘body’ of possible methods, models, and techniques are applied to treat with human error management, cognitive processes, and HSI design.

The major analytic processes of HERMES are described in the following [2]. (a) Evaluation of socio-technical context and theoretical stand. (b) Root cause analysis (RCA) and accident/incident investigation. (c) Task boundary and initial condition identification. (d) Unwanted consequences and hazards evaluation. (e) Erroneous modes identification. Performance influencing factors (PIFs), cognitive functions, and error mechanisms should be used to evaluate the actual performance of the target task [1]. (g) Assessment of human error probabilities. The HERMES methodology has shown efficiency and effectiveness in a real and complex application [2]. The most important characteristic of HERMES model is to present its framework based on the consistent and integrated application of human factors methods and approaches. The implementation of HERMES is presented for two areas: first, the performance of an extended HRA in the framework of a quantitative risk assessment; and second, the implementation of safety assessment in a traffic organization.

3 Critical Elements of Human Reliability Assessment Techniques

3.1 Task Analysis

Task analysis is a fundamental approach for the HF expert. Task analysis refers to methods of properly describing and analyzing human systems interactions. Before HRA assessors begin to consider what errors might occur, they should have a clearly definition of what the operator should do to achieve correct performance. Although

there are many different ways of task analysis (TA) techniques, they all follow the same basic philosophy of evaluation, such as problem-perspective task definition, TA tool selection and application, task representation, and comparisons of actual and ideal task.

3.2 Human Error Identification

Taylor-Adams and Kirwan [22] concluded that a human error can be classified into four major components, including external error mechanisms (EEMs), internal error mechanisms (IEMs), performance shaping factors (PSFs), and psychological error mechanisms (PEMs). EEMs refer to the consequences or observable manifestation of the error, i.e. 'what error occurred'. For example, "valve left open," associated with each operator error, can also be determined in some but not all cases [22]. PEMs describe how the error occurred in terms of the psychological mechanism, e.g., cognitive tendencies and biases, and information processing capacities [6]. IEMs describe the internal manifestation of the error within each cognitive domain, e.g., late detection, misidentification, and hearback error [6]. HEI techniques deal with the question of what can go wrong in a system from human error perspective.

3.3 Human Error Representation

FTA, as defined in NUREG/CR-4835 [4], is an analytic technique used to find all possible situations that a system can fail. FTA is a graphically representative model of all the parallel and sequential combinations of faults that result in a predefined undesired event. Logic gates are fundamental to fault tree logic. The OR gate refers to a situation where the output event exists if any of the events under the OR gates exists. The AND gate refers to a logical operation where events under the AND gate must occur in order to produce the event. HRA assessors must use care in determining whether events related to human system interface fall within an AND or OR gate logic. An example was described in Figure 1 [4].

HRA event trees demonstrate operator actions in response to an initiating event and the successful or failed actions associated with the normal conduct of operation. HRA event trees also provide the information about recovery actions by operators or their supervisors and allow for modeling of errors of omission and commission. The best description of this method for representing human activity is found in NUREG/CR-1278 [21]. Information for developing the HRA event trees is determined from task analyses, structured interviews with a sample of plant personnel, human-machine systems analyses, review of piping and instrumentation diagrams, procedures, operating schematics, licensee event reports (LERs), and recent plant events. The left path represents the desired performance, and the right path represents deviation from that desired performance, shown in Figure 2 [21]. The limbs of the tree are assigned values, which may be taken from the technique for human error rate prediction (THERP) or elsewhere.

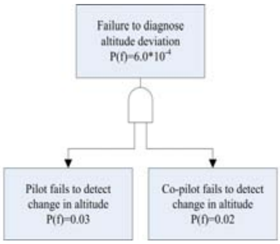


Fig. 1. An example of FTA (adapted from Haney et al. [4])

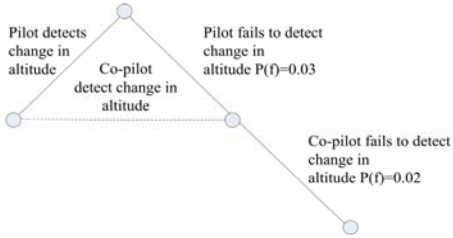


Fig. 2. An example of ETA (adapted from Swain and Guttman [21])

3.4 Human Reliability Quantification

Human reliability quantification techniques all involve the calculation of the human error probability (HEP), which is a measure of human reliability assessments. HEP is defined as follows [8]:

$$\text{HEP} = \frac{\text{The number of times an error has occurred}}{\text{The number of opportunities for that error to occur}} \tag{1}$$

Taylor-Adams and Kirwan [22] described that there are a wide variety of human-error probability data types: real operating experience, simulator data, experimental (performance literature) data, expert judgment, and synthetic data (i.e. from human reliability quantification techniques). These data types can originate from various data sources, such as: incident and accident reports, maintenance report, PSA reports, equipment records, interviews with plant personnel, near-miss reports, violation, plant log books, simulators, and experts [22]. In reality, Taylor-Adams and Kirwan [22] proposed that HEP is difficult to be gathered. There are four reasons to explain the lack of available of such data, that is, (a) difficulties involved in estimating the number of opportunities for errors in realistically complex tasks (the so called denominator problem), (b) confidentiality, (c) an unwillingness to publish data on poor performance, (d) a lack of awareness of why it would be useful to collect such data in the first place, and hence a lack of financial resources for such data collection.

3.5 Impact Assessment and Error Reduction Analysis

(1) Impact assessment

The impact assessment is an iterative process. Initially, if the risk level is too high, or incorrectly close to the criteria levels, the risk analysis will carry out an initial exploration of ways to reduce the risk level. The investigation will obviously focus on major risk-generating events. On the hardware aspect of PSA analysis, ways may be sought to improve reliability, e.g., by increasing the number of back-up (redundant) machines, by increasing the diversity of equipment used, or by using a more reliable component [8].

(2) Error reduction analysis

Error reduction analysis has already been mentioned in HRA process, including the following items [8]. (a) The use of task analysis to identify and reduce errors. (b) The use of human identification methods to identify errors and derive error

reduction approaches. (c) The use of PSA sensitivity analysis methods to identify ‘sensitivity’ errors which can be targeted for error reduction. (d) The use of quantification methods with built-in error reduction strategies. (e) The use of quantification methods with error reduction analysis capabilities.

4 Implementation of HRA in HSI Design Review

Computer based HSI is emerging as part of the new design of human machine systems. The impact of advanced HSI on the operator performance and plant safety should be evaluated before adopting them in the plants. In nuclear industry, advanced control room and human-machine interfaces are developed with advanced instrumentation and control (I&C) based on digital technology. There are complex operator aid systems being utilized based on modern computer technology, e.g., the computerized alarm system or large display panel [7].

4.1 The Main Purposes of Implementing HRA on HSI Design

The HRA methods have been investigated as a design tool and part of probabilistic safety analysis (PSA) that assesses various risks associated with complex systems [12]. Therefore, HRA can be used as an evaluation tool to identify human errors and system failures of the advanced HSI and control room. How the HRA methods can be implemented for an integrated validation of advanced HSI are summarized below [7].

First, the design objectives of HSI from a standpoint of human reliability are described as follows: (a) Maximization of human performance capabilities. (b) Minimization of human error probabilities. (c) Maximization of recovery possibilities of human errors. Second, in order to evaluate the HSI design, all the factors that may influence the human performance or human error probabilities should be considered together. In addition, what impact the availability of an operator aid system in the control room makes on the human reliability also should be estimated for the HSI design review.

4.2 The Most Discussed Issues of Implementing HRA on HSI Design

Human reliability has important implications not only in HSI design, but also in how human operators are used in human-machine systems. Considering the factors which decrease human reliability can be clearly identified. These factors are describes as follows [15]. (a) Training. Much of the effort in human reliability and system safety has been concentrated on personnel selection, placement, and altering the human by training. (b) Function allocation between humans and machines. It is necessary to identify what functions of HSI should be assigned to humans or machines. (c) Task analysis. As machines can be designed to fit human capabilities and limitations, tasks can also be designed for humans. (d) Anthropometric considerations. Anthropometry is the science of measurement and the technique of application that establishes the physical geometry, mass properties, and strength capabilities of the human body [17]. (e) Human factors checklists. Sanders and McCormick [18] proposed that using a checklist can be a systematic way to perform HFE evaluation.

4.3 Future Developments of Hra on High Hazard Hsi Design

By developing an understanding of the causes, modes, and probabilities of human errors, the HRA can provide valuable insights into the desirable characteristics of the design. Consequently, special attention should be paid to those scenarios, human actions, and HFE components that were identified by HRA and PSA analyses as they are important to the plant's safety and reliability. The future developments of HRA on high hazard HSI design are summarized bellow in three aspects.

(1) Aspects of human information processing model
The human information processing (HIP) model includes the perception, sensory storage, working memory, long-term memory, decision making, and attention. Human actions and errors can be thought to occur from the course of HIP. The difference between human actions and errors are that human actions can achieve the correct result, but the human errors result from the wrong actions. Therefore, paying attention to HIP model should be done before reducing human errors in high hazard design.

(2) Aspects of fuzzy set theory
Richei et al. [16] stated that most of the well-known procedures used for the probabilistic assessment of human factor events all include the use of vague linguistic statements on PSF to choose and to adapt fundamental human error probabilities from the associated databases. To improve the precision of human error probabilities, fuzzy set theory (FST) can be helpful to modify the value of human error rates estimation.

(3) Aspects of human errors checklist
For high hazard HSI designers and assessors, it is very convenient to use a systematic human error checklist. Thus, designing a systematic human error checklist for high hazard HSI is valuable. However, how to define the principles in developing the human error checklist appropriately is a very important issue. To clearly identify the differences between human factors checklist and human error checklists is also a critical issue.

5 Conclusions

Human reliability assessment in high hazard HSI design is a valuable activity which requires the integration of five components: (a) task analysis, (b) human error identification, (c) human error representation, (d) human reliability quantification, (e) impact assessment and error reduction analysis. A great deal of effort has been made on these components in the literature. What seems to be lacking, however, are the issues of methods review and future developments on high hazard HSI design.

This paper is intended as a review of the well-known HRA methods, which includes THERP, HEART, CREAM, HEROS, and HERMES. These HRA methods have similar properties, such as the analytic target, generic process, applicable scope and limitations. However, there still are some characteristics which are quite different from each other. For example, the knowledge base of HEROS is derived from the results of the generally valid ergonomic and psychological studies. These results are

expressed by fuzzy variables. Therefore, it is not necessary to use existing databases, e.g., human error databases of THERP.

There are advantages and disadvantages in each HRA methods. HRA assessors need to know how to use these methods at the right situation. This paper provides a categorization of HSI that HRA assessors can learn from the property of current HRA methods utilized in high hazard human-system interface design. Finally, this paper provides three future development needs of HRA on high hazard HSI design. They are aspects of human information processing model, aspects of fuzzy set theory, and aspects of human errors checklist. These future trend viewpoints of HRA methods should be further developed for new HRA methods on high hazard HSI design.

Acknowledgments. The preparation is partially funded by study is supported by Institute of Nuclear Energy Research (962001INER008).

References

1. Cacciabue, P.C.: Modelling and simulation of human behaviour in system control. Springer, London, UK (1998)
2. Cacciabue, P.C.: Human error risk management for engineering systems: a methodology for design, safety assessment, accident investigating and training. Reliability engineering and system safety 83, 229–240 (2004)
3. Döös, M., Backström, T., Sundström-Frisk, C.: Human actions and errors in risk handling—an empirically grounded discussion of cognitive action-regulation levels. Safety Science 42, 185–204 (2004)
4. Haney, L.N., et al.: Comparison and application of quantitative human reliability analysis methods for the risk methods integration and evaluation program (RMIEP), NUREG/CR-4835, U. S. Nuclear regulatory commission, Washington, DC (1989)
5. Hollnagel, E.: Cognitive Reliability and Error Analysis Method—CREAM, 1st edn. Elsevier Science, Oxford, England (1998)
6. Isaac, A., Shorrock, S.T., Kirwan, B.: Human error in European air traffic management: the HERA project. Reliability engineering and system safety 75, 252–272 (2002)
7. Kim, I.S.: Human reliability analysis in the man-machine interface design review. Annals of nuclear energy 28, 1069–1081 (2001)
8. Kirwan, B.: A guide to practical human reliability assessment. Taylor & Francis, London, UK (1994)
9. Kirwan, B.: The validation of three Human Reliability Quantification techniques—THERP, HEART and JHEDI: Part 1—technique descriptions and validation issues. Applied Ergonomics 27(6), 359–373 (1996)
10. Kirwan, B.: Human error identification techniques for risk assessment of high risk systems—Part 1: review and evaluation of techniques. Applied ergonomics 29(3), 157–177 (1998(a))
11. Kirwan, B.: Human error identification techniques for risk assessment of high risk systems—Part 2: towards a framework approach. Applied ergonomics 29(5), 299–318 (1998(b))
12. Modarres, M., Kaminskiy, M., Krivtsov, V.: Reliability engineering and risk analysis: a practical guide. Marcel Dekker, New York, USA (1999)
13. Mosleh, A., Chang, Y.H.: Model-based human reliability analysis: prospects and requirements. Reliability engineering and system safety 83, 241–253 (2004)

14. Munger, S.W., Smith, R.W., Payne, D.: An index of electronic equipment operability, data store, AIR-C43-1/62-RP (1). America Institute for research, Pittsburgh (1962)
15. National Safety Council (NSC). Accident prevention manual for industrial operations. National Safety Council, Itasca, USA (1974)
16. Richei, A., Hauptmanns, U., Unger, H.: The human error rate assessment and optimizing system HEROS—a new procedure for evaluating and optimizing the man-machine interface in PSA. Reliability Engineering and System Safety 72, 153–164 (2001)
17. Roebuck, J.A.: Anthropometric Methods: Designing to Fit the Human Body. Human Factors and Ergonomics Society, Santa Monica, CA (1995)
18. Sanders, M.S., McCormick, E.J.: Human factors in engineering and design. McGraw-Hill, Inc, Singapore (1992)
19. Stanton, N., Baber, C.: A system approach to human error identification. Safety science 22(1-3), 215–228 (1996)
20. Swain, A.D.: THERP, SC-R-64-1338. Sandia National Laboratories (1964)
21. Swain, A.D., Guttmann, H.E.: A handbook of human reliability analysis with emphasis on nuclear power plant applications. NUREG/CR-1278, USNRC, Washington DC., USA (1983)
22. Taylor-Adams, S., Kirwan, B.: Human reliability data requirement. The International Journal of Quality & Reliability Management 12(1), 24–52 (1995)
23. Williams, J.C.:HEART – a proposed method for assessing and reducing human error. Ninth Advances in Reliablity Technology Symposium, NEC, Birmingham, June, AEA Technology, Culcheth, Warrington (1986)

Who Is Taking over Control? A Psychological Perspective in Examining Effects of Agent-Based Negotiation Support Technologies

Yinping Yang[1], John Lim[1], Yingqin Zhong[1], Xiaojia Guo[1], and Xue Li[1]

[1] School of Computing, National University of Singapore, 3 Science Drive 2, Singapore 117543
{disyyp, jlim, diszyq, disgxj}@nus.edu.sg, lixuer@gmail.com

Abstract. This paper attempts to understand the effects of agent-based negotiation mechanisms through human's psychological perspective. We argue that the impact of automated negotiation, despite its efficacy and effectiveness, may result in a shift of negotiators' beliefs on their control over the choice of negotiation tactics and decision making process; such loss of control in turn causes an increase of user anxiety towards the system. In addition, individual differences, such as negotiators' decision-making style patterns are posited to moderate the relationship of system types and users' perceptions. The study involves an experiment employing randomized block design. Findings suggest significant impact of types of negotiation support systems on perceived control, as well as a negative relationship between perceived control and system anxiety. However, the moderating effect of decision-making style patterns is not evident in our data. Discussions and implications are drawn.

Keywords: perceived control, system anxiety, negotiation support systems, intelligent agents, decision-making style.

1 Introduction

Negotiation has long been a subject of study by economists, psychologists and organization theorists. During the last three decades, there has been an evident interest in applying computer technologies to solve problems encountered in negotiations; such computer-based systems which typically involve group and decision support modules, communication tools, and/or agent-based techniques, are collectively known as Negotiation Support Systems (NSS). The empirical research on traditional NSS has generally proved that NSS help to improve the negotiation outcome in terms of higher joint utility and greater satisfaction [7, 12, 20]. More recently, intelligent software agents open up new and exciting possibilities for automating business negotiation in electronic marketplace. The use of negotiation agents are potentially advantageous in saving human negotiators' time in monitoring market change, avoiding unnecessary cognitive conflict like "face-saving behavior", and lowering transaction cost [18].

Despite the growing interest in the design and implementation of negotiation agents (e.g., [5, 16, 18, 23]), little has been done to empirically investigate the impact

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 222–231, 2007.

of various agent-based technologies on the psychological changes of human negotiators. Comparing to the earlier emphasis in developing systems to support or *supplement* negotiators' decision making, in the newer approach, agents who negotiate on behalf of their owners are designed to effectively *supplant* negotiators in reaching settlements [3]. An important notion here is the degree of *control* which may be perceived differently by human negotiators. Following our earlier works [26, 27], this study is designed to examine the effects of various negotiation support techniques on the key psychological constructs of users' perceived control and system anxiety. Additionally, negotiators' decision-making style patterns are posited as moderating factors affecting the above relationship.

We first provide the theoretical foundation on negotiation support systems according to the "types of NSS", followed by user-related aspects on perceived control and decision-making style model, and their influences on system anxiety. Research model and hypotheses are derived based on the theoretical arguments. The following sections outline the research method, and present the results of data analysis. Discussions and implications of the findings are addressed.

2 Theoretical Foundation and Model

2.1 Types of Negotiation Support Systems

Early design on NSS aims to provide computer-based support for face-to-face negotiations and typically incorporates two major components – the individual Decision Support System (DSS) for each party and the electronic communication channel [14]. The DSS portion improves the human information-processing capacity and its effects can be best understood within the reference disciplines of game theory and economic theory. The electronic communication channel improves perceived commitment which can be best understood within the context of social-psychological theories of negotiation. In empirical research, several NSS system prototypes are developed reflecting the two major functions. For example, the series of studies [7, 9, 12] include a module for electronic communication between the parties, and a DSS to assist in the generation and evaluation of alternative contracts. Each negotiator provides his/her own interests to the decision tool as private data, as well as his/her guesses of the other party's preferences. The DSS then computes and displays the three contract alternatives with the highest joint outcome. The electronic communication module is integrated to facilitate text-based message and file transfers between negotiation parties. Collectively, this type of NSS is labeled as in our study *Type I System with Decision Support Focus.*

The type II and type III systems deal with agent-based negotiation support technologies. Generally, an automated negotiation involves two or more electronic agents supporting negotiation parties in a virtual environment governed by computational rules [9]. The design of autonomous negotiation aims to fully automate the negotiation process by implementing self-interested electronic agents. By utilizing concession-based negotiation tactic [15], Agents can make small concessions in each proposal without losing patience as human beings may do. Such agent-based system is very straightforward for negotiators to understand and easy to use, which would

greatly ease negotiators' effort and time spent in the negotiation process (e.g., [9]). This kind of design is labeled as *Type II System with Automated Negotiation Agent.*

Although being easy to understand and use, the relatively simple concession-based negotiation agents have limited capabilities in representing the social-psychological complexity of human negotiators. Recently, AI researchers advocate and examine the use of more intelligent agents that can provide personality features and learning capabilities to capture inputs of human negotiators [10, 16, 28]. To incorporate people's different "personalities" into a negotiation agent, the user can plug in suitable "subjective beliefs" to his or her agents. With different "internal beliefs", agents can behave quite differently to imitate human's behaviors: some are tough negotiators, while others are accommodating people who are very willing to make a deal with their opponents. Huang and Sycara [10] further present a computation model of such personalized agent-based negotiations where the participating agents with "internal beliefs updating methods" (or learning features) change their behaviors in accordance to adjustments by users of some critical parameters. In every negotiation iteration, an agent checks the history of the process, updates its beliefs about its opponents, and then tries to maximize its own expected payoff based on its own subjective beliefs. It has been shown that under qualified conditions, software agents with learning heuristic can act as effective surrogates of human negotiators [5]. Arguably, the *Type III: Automated Negotiation Agent with Personality and Learning Features* is the most advantageous design for negotiators as it benefits from the agents' computational capacities and incorporates learning capabilities and adaptive strategies that can be updated by human negotiators.

2.2 Perceived Control

The three alternate system designs outlined above differ significantly in terms of users' involvement into the negotiation process. An interesting issue that deserves further examination is how the types of NSS differ in influencing users' perceptions towards their ability to control. The construct "perceived control" comes from social psychology research. Perceived control is viewed as the degree to which a person feels that he/she can impact outcomes in his/her environment through voluntary actions [13]. An individual's perception on the control that he can exert has been found to be a very strong predictor of both behaviors and emotional outcomes [13], and therefore has stimulated a great deal of research in disciplines such as psychology, marketing, and organizational behavior. Averill [1] views it as a multi-dimensional construct which include cognitive control, behavioral control, and decisional control. Cognitive control addresses the interpretation of an event into a cognitive model or plan. Decisional control addresses the ability to choose among different courses of action. Behavioral control deals with the existence of some means to exert influence over an event. This dimension of control has been extensively researched involving Azjen's [2] theory of planned behavior.

In research on computer information systems, the notion of perceived control is adopted to understand human's perception towards system use. However, the control construct is not examined using multiple dimensions (e.g., [24]). Frese [8] integrates the various aspects of control into the information system domain; it provides conceptual discourse on the aspects of control form an information systems

perspective. Based on this concept, Morris and Marshall [17] developed and tested 55 items concerning user's experiences working with an interactive information system. The factor analysis produced five key factors that represent perceived control: feedback signal, feedback duration, strategy, metaphor knowledge, and timeframe.

In our study, we adopt the three-dimensional concept of *perceived control* to assess the degree to which negotiators are able to voluntarily utilize system to achieve their expected outcome. If trainings are provided for understanding the functionality of all the three types of system, cognitive control may not vary significantly from one another. Hence, perceived control may mainly differ in the aspect of decision making and action taking. Fully automated arbitration through negotiation agents would constitute an extreme case of a change in the locus of control from humans to computer agents [20]. Type I and II systems are presenting the two extreme cases. The reason is that NSS with decision support focus on mainly providing supplementary analytical tools. They still require human negotiators to make decision and take the action to negotiate with opponents (decision and behavioral control); while agent-based NSS without personality on behalf of the human users take over the control on both action taking and decision making. Type III system stands in between, because users can plug in their preferences and adjust agent's negotiation strategy in any time during the negotiation process (behavioral control).

H1a. NSS with decision support focus (type I) will result in higher perceived control than agent-based NSS without personality and learning features (type II).

H1b. Agent-based NSS with personality and learning features (type III) will result in higher perceived control than agent-based NSS without personality and learning features (type II).

2.3 Decision-Making Style Patterns

Negotiators may have different styles and preferences to approach the decision making process. By understanding the decision style one may be able to predict outcomes in term of decision behavior [21].

There are different definitions of decision making style in the literature. One widely accepted model [21, 22] compromises two dimensions: cognitive complexity and values orientation. *Cognitive complexity* is the measure of tolerance for ambiguity. People with high tolerance for ambiguity are able to cope with a high degree of uncertainty, and are more idea-oriented than action-oriented. This personality leads to the desire for more information from multiple sources and consideration of many alternatives [21]. NSS with decision support focus are designed to generate optimal solutions (e.g., Alternative Generator function computes a range of choice suggesting contract alternatives leading to highest joint utilities). In contrast, fully automated agent-based NSS take the role of negotiating on behalf of negotiators, based on built-in algorithms with which details are not (and unnecessarily) explicit to users. In this situation, cognitively-complex people who are willing to solve problems by themselves may feel losing control; in other words, the more automated the process can be provided by the system, the stronger of the feeling is towards losing control.

As all the three systems provide ways to structure the negotiation process, cognitively-structured people who tend to avoid ambiguity and prefer structure and

specific instructions, will perceive little difference of control towards the three types of systems.

H2a. Cognitively-complex negotiators will experience highest perceived control towards NSS with decision support focus (type I) followed by agent-based NSS with personality and learning features (type III) and lowest over agent-based NSS without personality and learning features (type II).

H2b. For negotiators who are cognitively-structured, there will be no significant difference in their perceived control towards the three types NSS.

From the *values orientation*'s dimension, task-oriented people who are characterized by analytical and directive styles generally have a need for tight control; in contrast, people-oriented are willing to compromise and accept loose control [21]. Agent-based NSS with personality and learning features can perform better through concession-based tactic and simulation of real-life negotiator's approach - experienced human negotiators can effectively learn from their opponents and estimate the opponents' preferences to find appropriate strategies during the communication process. In this situation, task-oriented negotiators may perceive highest control since the type III system is designed towards achieving best performance in terms of outcome utility, and may perceive very low control towards type II NSS as they do not provide adaptable negotiation strategies.

People-oriented negotiators who care for effective communication may not feel too much difference from the three NSS, as the communication function is consistently available in all.

H2c. Task-oriented Negotiators will experience highest perceived control towards agent-based NSS with personality and learning features (type III) followed by NSS with decision support focus (type I) and lowest over agent-based NSS without personality and learning features (type II).

H2d. For negotiators which are people-oriented, there will be no significant difference in perceived control towards the three types NSS.

2.4 System Anxiety

The term anxiety is most often used to describe an unpleasant emotional state or condition which is characterized by subjective feelings of tension, apprehension, and worry [25]. In human-computer studies, a most widely studied aspect is computer anxiety, which can be defined in terms of a psychological response (e.g., computer phobia) [19] or in terms of a cognitive reaction (e.g., apprehension of computer technology) [11]. *System anxiety* which is considered as the anxiety associated with the particular use of a system comes under the umbrella of computer anxiety [4].

Since locus of control has been regarded as one salient correlate of computer anxiety in the literature (e.g., [6]), perceived control is conceivable to be relevant to users' anxiety towards a particular system. With agent-based NSS come into the business, there is a great change in the degree of control from humans to computer agents and may consequently result in higher system anxiety perceived by users. System anxiety may cause some negative behaviors like avoidance of NSS usage, negative comments about the system, and further influence the final outcome of

negotiation in an undesired way. We hypothesize users with higher perceived control are expected to experience lower anxiety towards using the system.

H3. The level of system anxiety is negatively related to the level of perceived control of the system.

In summary, the research model and hypotheses are depicted in Fig. 1.

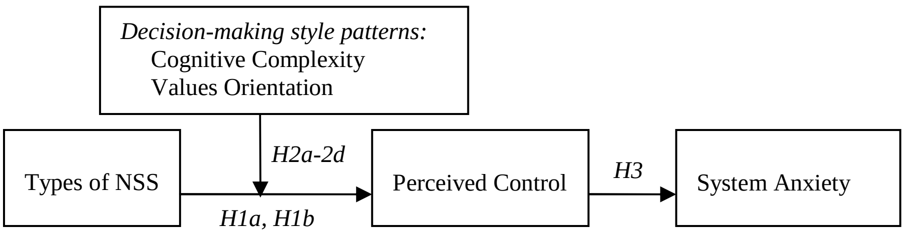


Fig. 1. The Research Model

3 Research Approach

An experiment was carried out to investigate the joint impact of the types of NSS and people's decision patterns. A Randomized Block Design (see Table 1) was employed. Level of conflict, as an important situational factor (task characteristics) that influences negotiation outcomes, was incorporated in the experiment design. Consistent with previous empirical work, two levels of buyer-seller conflict of interests were used [7, 9, 12]. Essentially, conflict level was controlled as a block variable for consistency with previous studies and minimizes the influence of task characteristics, which also enhances the generalizability of this study. Other variables such as age, gender, past experiences and computer efficacy were controlled through randomization. Negotiation power was controlled in assigning same BATNA (Best Alternative to a Negotiated Agreement) to buyer and seller.

Table 1. Experimental Design

Block	**Types of System**		
	Type I	Type II	Type III
Low conflict	20 subjects	18 subjects	18 subjects
High conflict	20 subjects	18 subjects	20 subjects
Total	*40 subjects*	*36 subjects*	*38 subjects*

The *types of NSS* were manipulated using three web-based NSS, called ProNeg. ProNeg is a collection of negotiation support system prototypes developed with various support features (see [26] for detailed descriptions on the system design). We captured the subjects' *decision-making style patterns* through pre-experiment questionnaires. These variables were measured by Decision Style Inventory (DSI) ([22]) which aims at testing the preferences when approaching a decision situation. In

order to manipulate two groups of people whose different patterns are diversified and even, the subjects whose cognitive complexity score ranging from 151 to 169 were not used for analysis in the testing on moderating effects. For similar reason, the data for subjects with task value score from 145 to 164 were disqualified and discarded. The items for *perceived control* are adapted from Morris and Marshall [17]'s 55-item questionnaire; we develop the measurement of perceived control towards NSS by breaking it down into three dimensions compromising cognitive control, behavioral control, and decisional control (see Table 2). *System anxiety* is the specific anxiety associated with the use of the NSS. We adopt an email system anxiety measurement (from [4]) which was adjusted to ProNeg focus (see Table 3).

Table 2. Items for Perceived control

Cognitive Control	PC_CC_1	It was easy for me to visualize the way in which ProNeg works.
	PC_CC_2	I understood, at least abstractly, the components of ProNeg and how they work.
	PC_CC_3	I have used other systems that are similar to ProNeg in terms of the functionality.
	PC_CC_4	The objectives I attempt on ProNeg are very similar to the ones that I have successfully completed in the past.
Decision Control	PC_DC_1	ProNeg allowed me to make decisions about the way I would achieve my expected outcome.
	PC_DC_2	ProNeg allowed me to devise my own strategies to complete my negotiation task.
	PC_DC_3	ProNeg allowed me to update my expected outcome during the course of negotiation.
	PC_DC_4	ProNeg allowed me to attempt to achieve my expected outcome in any manner I want.
Behavior Control	PC_BC_1	I was able to use a very flexible approach to complete my negotiation task using ProNeg.
	PC_BC_2	I easily adjusted my negotiation strategies when I felt necessary to do so.

Table 3. Items for System anxiety

System Anxiety	SA_1	Using ProNeg made me nervous.
	SA_2	Using ProNeg made me uneasy.
	SA_3	I felt comfortable using ProNeg. (R)
	SA_4	I felt tense using ProNeg.
	SA_5	I felt uncertain using ProNeg.

The *task* was adapted from Jones's work [12]; it involves a real-life multi-issue negotiation scenario in which buyers and sellers bargain for a purchase agreement for an engine sub-component. All negotiation sessions were conducted online where each buyer and seller pair was seated in different locations and communicated over web-based systems. This is to simulate e-negotiation activities across geographical span.

4 Results and Discussions

MANOVA test was carried out to test the hypotheses 1a and 1b[1]. Conflict level as a block variable was also employed to this test as a fixed factor to control its block effect. Age (detected to be affected by block variable in the control check) was employed as a covariate to remove its nascence effect over the independent variables. Results indicate that types of NSS have significant main effect on subjects' perceived control (F=8.486; P=0.001), but impact insignificantly in system anxiety (F=2.367; P=0.099). Contrast tests were conducted to further test the specified relationships in hypotheses 1a and 1b. The results (without assuming equal variances) suggest that H1a is supported (P=0.000) and H1b was not supported (P=0.170).

In line with *H1a*, NSS with decision support focus result in higher perceived control than agent-based NSS without personality features. This implies that total automation, despite its efficiency and effectiveness, is not the optimal design for NSS when people desire a self control for negotiation activities. From a human-centric perspective, being blocked away from the decision making process causes people to feel a loss of control. This is an important message to researchers interested in system design. Psychology research also suggests that control is the general motive for human activities; losing motive may result in a negative relation towards the final outcome. However, the data also reveal that users did not perceive higher control over NSS with personality and learning features than agent-based NSS without personality (*H1b*). One possible reason may lie in the negotiation task involving four negotiation issues which only generate a total of 728 contract alternatives. Arguably, this task might not be a very cognitively demanding, which causes the agent-based NSS to be under-appreciated by users. When users are confronted with a more complicated task involving more issues with tremendous cognitive efforts beyond the capability of human negotiators, people may better appreciate type III system.

Two-way MANOVA test showed no interaction effects between types of systems and decision-making style patterns on perceived control. *H2a* to *H2d* are not supported. A possible reason is related to the operationalization of unevenly distributed decision-making style patterns. Students may represent biased groups (inclined people-orientated and cognitively-structured styles based on our analysis) and such bias and difficulty to manipulate decision style patterns may in turn result in the insignificance of the asserted moderating effect. The second reason lies in the defined negotiation situation. The pre-defined goal and issues which aim to clear the task would have reduced the different actions and perceptions to be undergone by people in terms of different patterns. More in-depth research is warranted to demonstrate more effective ways to measure individual traits.

Linear regression test indicated that the level of anxiety is negatively related to perceived control (Beta = -0.467). *H3* positing the negative relationship between perceived control and system anxiety is supported, in line with the psychology literature. System anxiety is an important psychological attitude that would affect negotiators' performance including achieving win-win solutions and greater

[1] Prior to hypotheses testing, control and manipulation checks have been performed. Items PC_CC_3, PC_CC_4 and SA_3 did not satisfy the discriminant and convergent criteria and were dropped. Due to page constraint, the detailed measurement development and statistical analysis process are not articulated; they will be available upon request.

satisfaction. Better understanding and assessing this variable may help to diminish a negative effect caused by system.

Also, it has been shown that there is a significant effect between system intelligent levels and perceived control as predicted in H1a and H1b, but no significant effect between system types and system anxiety. Therefore, by definition, perceived control is a mediating variable between types of NSS and system anxiety. Overall, the model explains the effects of system artifacts over human perception during the negotiation process.

5 Concluding Remarks

The paper has emphasized and captured an empirical investigation of perceived control in assessing the effect of negotiating agents. It produces a parsimonious eight-item measurement from three dimensions of control identified by psychology literature; the measurement is expected to situate more properly perceived control in the realm of agent-based software research. With a better understanding on perceived control from users' perspective, the design and implementation of negotiation support tools can be directed to focus on the intrinsic needs of the negotiators. Furthermore, this study is one of the first to look into the impact of various forms of designs of negotiation support systems including autonomous agents. Instead of assuming an input-output approach, we attempt to elucidate the underlying mechanism on how negotiation process is perceived differently in using different types of support systems. The realization of mediating role of perceived control between different types of system and system anxiety in the negotiation process would contribute to a deeper understanding of the socio-emotional aspect of negotiation outcomes.

References

1. Averill, J.R.: Personal Control over Aversive Stimuli and Its Relationship to Stress. Psychological Bulletin 80, 286–303 (1973)
2. Azjen, I.: Theory of Planned Behavior. Orga.l Behav. and Human Deci. Process 50, 179–211 (1991)
3. Blecherman, B.: Adopting automated negotiation. Tech. in Society 21, 167–174 (1999)
4. Brown, S.A., Fuller, R.M., Vivian, C.: Who's Afraid of the Virtual World? Anxiety and Computer-Mediated Communication. J. of the Association for Information Systems 5, 79–107 (2004)
5. Chari, K., Agrawal, M.: Multi-issue Automated Negotiations Using Agents. INFORMS J. of Comp. (Accepted, Forthcoming, 2006)
6. Crable, E.A., Brodzinski, J.D., Scherer, R.F., Jones, P.D: The Impact of Cognitive Appraisal, Locus of Control and Level of Exposure on the Computer Anxiety of Novice Computer User. J. of Edu. Comp. Research 10, 329–340 (1994)
7. Delaney, M.M., Foroughi, A., Perkins, W.C.: An Empirical Study of the Efficacy of a Computerized Negotiation Support System (NSS). Deci. Sup. Sys. 20, 185–197 (1997)
8. Frese, M.: Theory of Control and Complexity: Implications for Software Design and Integration of Computer Systems into the Work Place. In: Frese, M., Ulich, E., Dzida, W. (eds.) Psychological Issues of Human Computer Interaction in the Work Place, Elsevier Science Publishers, Holland (1987)

9. Goh, K.Y., Teo, H.H., Wu, H.X., Wei, K.K.: Computer-Supported Negotiations: An Experimental Study of Bargaining in Electronic Commerce. In: Proc. of the 21st Annual Int. Conf. on Info. Sys. Brisbane, Australia (2000)
10. Huang, P., Sycara, K.: A Computational Model for Online Agent Negotiation. In: Proc.of the 35th Hawaii Int. Conf.on Sys Sci. Hawaii (2002)
11. Joncour, N., Sinclair, K., Bailey, M.: Computer Anxiety, Computer Experience and Self-Efficacy. In: Proc.of Annual Conf.of the Australian Assoc. for Research in Edu. Newcastle, New South Wales (1994)
12. Jones, B.H.: Analytical Negotiation: An Empirical Examination of the effects of Computer Support for Different Levels of Conflict in Two-Party Bargaining. In: Ph.D. Dissert. Indiana University, Bloomington, Indiana, USA (1988)
13. Lacey, H.M.: Control, perceived control, and the methodological role of cognitive constructs. In: Perlmuter, L.C., Monty, R.A. (eds.) Choice and Perceived Control, Lawrence Erbaumn Associates, Hillsdale, NJ (1979)
14. Lim, L.H., Benbasat, I.: A Theoretical Perspective of Negotiation Systems. J. of Manag. Info. Sys. 9, 27–44 (1993)
15. Lopes, F., Novais, A.Q., Mamede, N., Coelho, H.: Negotiation Tactics for Autonomous Agents. In: 12th Int. Workshop on Database and Expert Sys. App, 0708 (2001)
16. Luo, X., Jennings, N.R., Shadbolt, N.: Acquiring user tradeoff strategies and preferences for negotiation agents: A default-then-adjust method. Int. J. of Human-Comp. Studies 64, 304–321 (2006)
17. Morris, S.A., Marshall, T.E.: Perceived Control in Information Systems. J. of Org. and End User Comp 16, 38–56 (2004)
18. Oliver, J.R.: A Machine Learning Approach to Automated Negotiation and Prospects for Electronic Commerce. J. of Manag. Info. Sys. 13, 83–112 (1997)
19. Ramalingam, V., Weidenbeck, S.: Development and Validation of Scores on a Computer Programming Self-Efficacy Scale and Group Analyses of Novice Programmer Self-Efficacy. J. of Edu. Comp. Research 19, 367–381 (1998)
20. Rangaswamy, A., Shell, G.R.: Using Computers to Realize Joint Gains in Negotiations: Toward an Electronic Bargaining Table. Manag. Sci. 43, 1147–1163 (1997)
21. Rowe, A.J., Boulgarides, J.D.: Managerial Decision Making: A Guide to Successful Business Decisions. Macmillan Publishing Company, New York (1992)
22. Rowe, A.J., Mason, R.O.: Managing with Style. Jossey-Bass, San Francisco (1987)
23. Sandholm, T.: Automated Negotiation. Commu. of the ACM 42, 84–85 (1999)
24. Sengupta, K., Te’eni, D.: Cognitive Feedback in GDSS: Improving Control and Convergence. MIS Quart. 17, 87–109 (1993)
25. Spielberger, C.D.: Conceptual and Methodological Issues in Anxiety Research. In: Spielberger, C.D., Sarason, I.G. (eds.) Anxiety: Current Trends in Theory and Research (2), pp. 481–494. Academic Press, New York (1972)
26. Yang, Y.P., Zhong, Y., Guo, X.J., Lim, J.: An Empirical Study on Negotiators’ Perceived Control and System Anxiety towards Agent-based Negotiation Support Tool. In: Proc. of Grp. Dec. and Nego. 2005. Vienna, Austria, (July 10-13, 2005)
27. Yang, Y.P., Lim, J., Chen, Y.: Alternate Designs on Agent-Based e-Negotiation Systems: Concepts and Implementation. In: 16th Workshop on Info. Tech. and Sys. WITS2006, Milwaukee, Wisconsin, USA, (December 9-10, 2006)
28. Zeng, D., Sycara, K.: Bayesian Learning in Negotiation. Int. J. of Human- Comp. Studies 48(1), 125–141 (1998)

Adaptive User Interactive Sketching for Teaching Based on Pen Gesture

Haiyan Yang[1,2], Cuixia Ma[1], Dongxing Teng[1], Guozhong Dai[1], and Hongan Wang[1]

[1] Institute of Software, The Chinese Academy of Sciences
[2] Graduate University of Chinese Academy of Sciences
Beijing 100080, China
haiyan820@hotmail.com, cuixia@ios.cn

Abstract. In this paper, the pen-based adaptive interface architecture for geometric teaching is presented to facilitate natural interactive sketching for geometric graphics application. Based on user-centered design, it analyzes the intent-based manipulation in terms of context based on pen gesture. By providing adaptive constraint capture and dynamic interaction this architecture can improve operation efficiency of the interface and reduce user's mental workload. Furthermore, it provides an educational application for benefiting teaching experience by preventing users from being interrupted in class.

Keywords: sketch, pen gesture, adaptive interface, intent-based method.

1 Introduction

Natural interaction aims to make human-technology interaction as intuitive as human-human interaction [1] [4]. Pen-based interface mode is much more convenience and natural for users to express themselves between human and computer, which can provide the imprecise and quick sketching to implement geometry operation in terms of common skill, such as pen gesture [2] [3]. Users prefer using pen to operating mouse when doing sketch, and pen-based interface offers many performance advantages, nonetheless not all system design is necessarily best approached with a pen-based interface.

As we know, WIMP interactive mode which based on mouse and keyboard provides graphical user interface, it is much more natural and wieldy than command line interface. But with rapid development of computer technology, the software size becomes larger. The interface is more complex than ever before. A lot of professional software affords so many functions that can meets the needs of user. But at the same time, users have to concentrate their attention on the applications and to find target objects from so many menus and icons. It not only distracts users' attention, but also leads to more cognitive loading. Especially in current geometry systems, designers are bounded in the WIMP interactive mode. They have to provide precise information and operate them according to the ordered steps of menu and icon. Pen-based adaptive interface has many advantages. First it is practical, it is very easy to learn and present

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 232–241, 2007.

naturally and efficiently. Compared with menus and buttons, it is easier for users to remember and use these mutual commands by this way. Second, using pen gesture to manipulate geometry graphics matches the common skill of people. Thirdly, with emphasis on invisible interaction, context-awareness technology and experience capture and access, it will provide good experience and reference for the research of new interaction method.

In this paper, we also provide a geometry teaching system, integrated with functions of blackboard, powerpoint and geometry sketchpad which are popular used in today. We analyze different roles that teacher performs during teaching work, and design the system based on teacher-centered. It affords different functions in distinct occasions, like that in the process of preparing for lessons or conducting prelection, to reduce mental workload and improve the efficiency in teaching.

2 Related Work

Geometer's sketchpad was provided by the company of Key Curriculum Press in USA. It is one kind of popular dynamic geometry software for teaching, which breaks the limitation of traditional geometry tool. Plane geometry is one of the Z+Z intelligent education platform soft wares. It integrates dynamic geometry with automatic reasoning. Besides the advantages of dynamic graphics of geometer's sketchpad, it is more convenient for teaching in intelligent drawing and automatic reasoning. Cinderella is also one kind of dynamic geometry software, which is developed by Jurgen Richert-Gebert and Ulrich kortenmp [5]. It provides many geometry entities and tools to create interactive graphics on web. It supports the function of theorem checking. The software is suitable not only for students to study the knowledge, such as Euclid geometry, ellipse geometry and hyperbola geometry and so on, but also for researcher to work on it. Also, some web designers can use it to implement some complex geometry graphics. There are also many related research results about the geometry software, such like the system of JavaSketchPad and cabri Java. Or some certain soft wares, space analytic geometry and chemical molecule geometry are developed by NIMA software Group.

Those soft wares above are all based on WIMP interaction paradigm, and are often more of a hindrance than a benefit during the process of drawing. These tools take too much time to use and force designers to specify more of the design details based on WIMP user interface. Additionally, those can be used as tool for geometric work, but actually not suitable for teaching. The university of Washington was developed and deployed Classroom Presenter, a Tablet PC-based presentation system enables an active lecturing environment [10]. The university of San Diego developed the ubiquitous presenter to expand presenter via common web technologies to support non-Tablet audiences and enhance student control [12]. Evan Golub introduced a computer-based presentation system modeled on handwritten transparencies [11]. Some tools are good at presenting and interacting in the classroom, but are short of rich flexible graphic operations.

In this paper, we introduce one system for geometry teaching and provide modification operations with the pen and adaptive interface, as is suitable to the common skill of users and to making the teaching work flexible.

3 Interactive Sketching for Teaching

3.1 Adaptive Constraint Capture

With the development of ubiquitous computing, many new methods and technologies are presented to improve the drawing process, such like natural interaction, context-awareness and experience capture and access. Natural interaction is convenient for users to operate the process freely instead of paying much attention to the tool itself. Especially in some certain applications, like geometry teaching, natural interaction is efficient and necessary. We give two kinds of interaction tasks for this process: sketch task and context-aware task. Sketch task refers to pen gesture commands; the context-awareness refers to users' intending operations. Context-aware condition is captured according to existing objects and relative behaviors, such as object type, operating position and so on. Context-awareness is the integral part of interaction computing [6].

Context refers to information which can show the entity states. The entities are users, location and objects. The context based on gestures in pen interaction is an instance of general context in pen interface [7]. The context can be divided into static and dynamic ones according to their state. The contexts such as object, constraint, current gesture are static. The contexts, such as user, location, scope which are related with specific operation, are dynamic. Generally, it is difficult to provide the clear interpretation based on the limited information from a sole context. The fusion of different context can provide more information (Fig. 1).

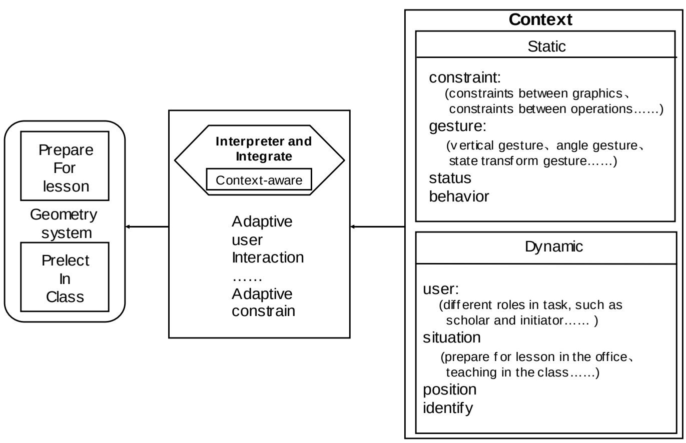
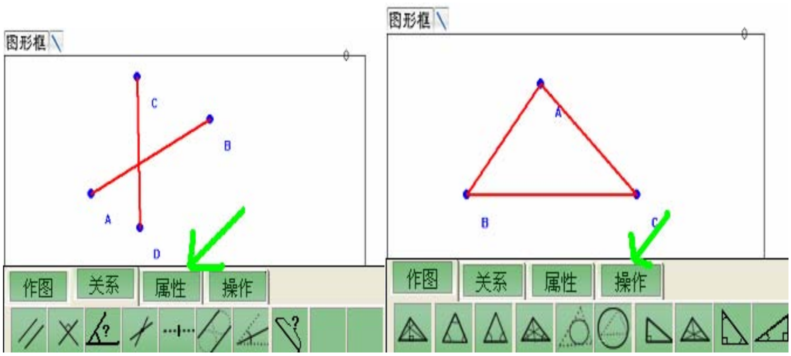


Fig. 1. Context-awareness model

We give an example to illustrate the constraint capture. If the user selects two lines, the system affords many operations relating to two lines according to the contexts (Fig. 2a). User could implement one operation without any confusion. Similarly, When the user selects a triangle, the system will offer operations relating to triangle, such as adding median of the triangle, changing it to equilateral triangle or isosceles triangle and so on

(Fig. 2b). During the process, constraints can be captured, and be kept in later operations. By simplifying the working environment and reducing the target task sets, user could focus his attention on the most important tasks, without being interrupted during the process of using the applications.



(a) adaptive constraint capture (b) another adaptive constraint capture

Fig. 2. Constraint capture based on user's intent

Teacher always have two main tasks during the teaching, preparing for lesson and prelection in the class. They are different tasks but closely connected. Geometry teacher always use three kinds of tools in the process, blackboard, powerpoint and geometry sketchpad. Teachers always use powerpoint to prepare the lessons and present materials in the class to reduce the time of writing on blackboard. Of course, Blackboard is also necessary for teacher to show dynamic process of teaching. Geometry sketchpad is an auxiliary tool for geometry teaching, and with it, teachers could draw graphics quickly and easily to show graphic dynamic track in the class. Both of them play an important role in the teaching, but each has its weakness. PowerPoint file is not easy to be modified in the class; while, to write down all contents on the blackboard is a waste of time; and current geometry sketchpad is conducted based on WIMP interaction paradigm, it is not natural and flexible for graphic manipulation. Teachers have to switch from one instrument to another frequently during the teaching process. It distracts teacher's attention and increases teachers' mental workload. Sometimes, it will disrupt user's task performance and emotional state.

We provide a system to afford a flexible way for geometry teaching. Teachers can concentrate their attention on the teaching work without any interruption. When teachers prepare for the lessons, the system provides several frames for various needs. For example, graphic frame is for geometry use; text frame can recognize free handwriting as formal text. We also provide picture frame, arithmetic frame, multimedia frame and so on (Fig. 3). User could create any type of frame at any place of the paper any time, without finding target commands in certain position of the interface.

Fig. 3. Various frames

During the prelection process, teachers emphasize on the interaction with students, and tasks are different from those of preparing for lessons. Referring to this character, the system can not only present teaching materials, but also provide operations for writing on blackboard and notation. Teacher could modify the objects at any moment, without returning to the preparing scene. In additional, we provide various tools for different use, such as fluorescent marker, palette and so on (Fig. 4).

Fig. 4. Auxiliary tools

3.2 Dynamic Interaction

User can sketch freely, just like writing on the normal physical paper, and the system could recognize rough handwriting to formal graphics. In additional, system can speculate the users' intent intelligently in terms of context-aware information. User could focus his attention on real tasks rather than unnecessary operations. There are two states during the process, drawing state and operating state. In drawing state, user can draw any rough sketch. System recognizes free handwriting to formal graphic synchronously. In operating state, system recognizes gestures to implement the appropriate operating based on user's intent captured. The switch between two states depends on a click gesture. We define a time threshold for click gesture. If the time of click operation exceeds the given threshold, system will change the state from one to another. Otherwise, it will be regards as normal click operation. In figure 5, we give the architecture of intent capture.

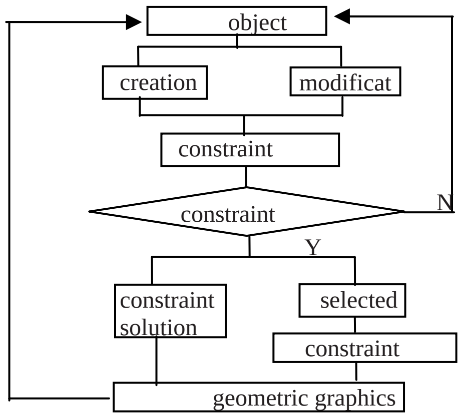


Fig. 5. Intent capture

The intent-based manipulation model is formally described as following: V is the geometry object; C is a set of constrains between the objects. The intention is expressed by a graph, V and C are two kinds of nodes, which are object node and constrain node respectively. There are three kinds of relations. The first one is the consequence of the objects, $\Gamma_{\text{int}}(v)$,$v \in V$; the second one is the relation between object and corresponding constrains, $\text{H}_{\text{int}}(v,c)$, $v \in V$, $c \in C$; the third one is the connection relation among the constrains according to the objects , $\text{K}_{\text{int}}(c)$, $c \in C$.

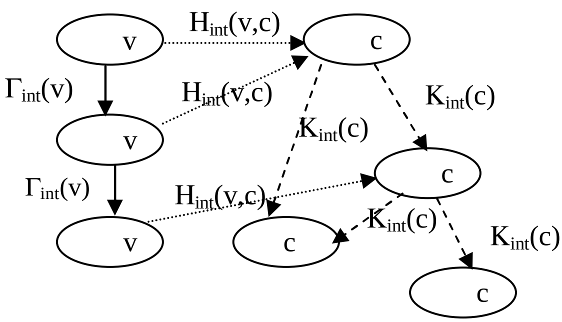
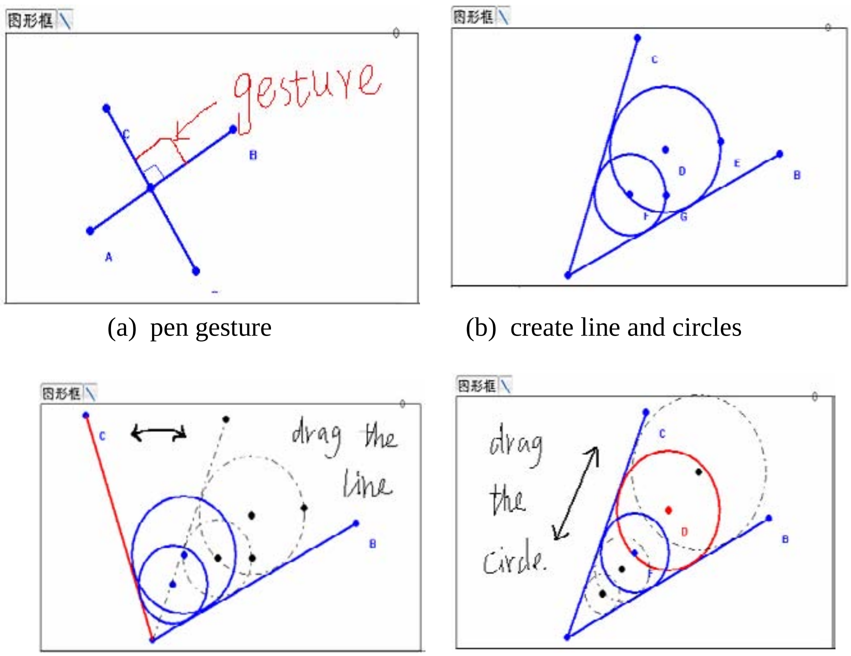


Fig. 6. Constraint architecture

When user presents a vertical gesture between two intersecting lines, the perpendicular constraint can be captured (Fig. 7a). User draws an angle and two circles, and adds tangency constraint between the circle and the angle (Fig. 7b). We can drag a line freely, and the constraints with circles are kept when the line is moving (Fig. 7c). Similarly, the constraint will be kept when we move the circles (Fig. 7d).



(a) pen gesture (b) create line and circles

(c) drag the line (d) drag the circle

Fig. 7. Dynamic constraint capture

4 System Architecture

The adaptive interface architecture for teaching provides flexible way for modeling interactive sketching for geometry operation. There are three important modules in the architecture, the support module, the interaction module and the task module. The support module involves the domain knowledge used in the process of interaction and recognition, including geometry knowledge base and gesture knowledge base. The interaction module deals with the command parsing of gesture and intelligent speculation, including context-aware and constraint, dynamic interaction and intent capture. The task module integrates all the common tasks in geometry teaching, and provides different interaction mode for different situation.

An architecture is given for geometry teaching. Teachers can interact with system naturally without any interruption, and concentrate their attention on the task. In this system, three layers are given according to the granularity: paper, frame and the content. Each of them can be regard as an agent, paper containing frame and frame containing the content.

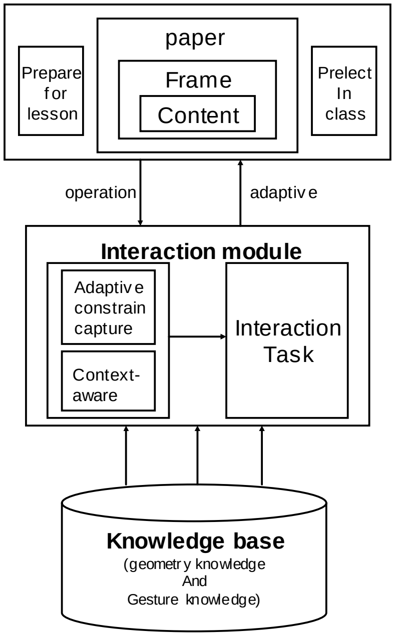
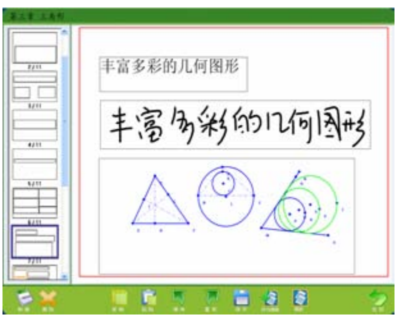
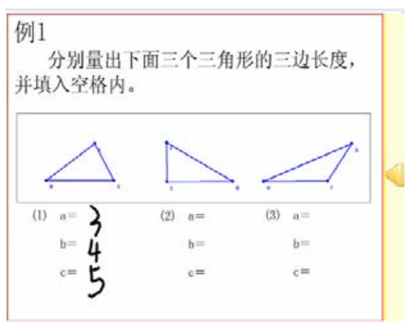


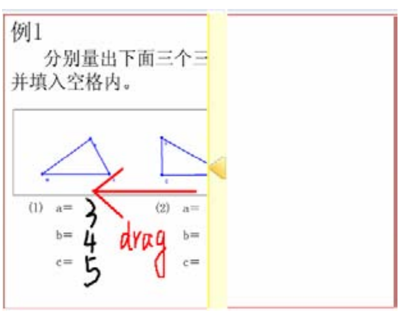
Fig. 8. System architecture



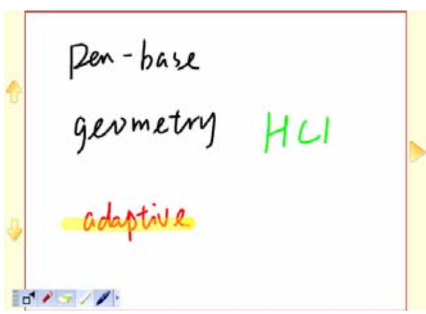
(a) preparing for lesson



(b) prelection and free sketch



(c) pull out whiteboard



(d) write on whiteboard

Fig. 9. The interface and some patterns

In additional, we also provide many auxiliary tools for ease of use, such as whiteboard functions, various types of pen, multimedia and so on. Whiteboard could provide a large blank area for teachers to do sketching, and it is very easy and natural for teachers to operate. Teachers only need drag the symbol from right to left to pull out the whiteboard or from left to right to hide it (Fig. 9c).

5 Future Works and Conclusion

Pen-based adaptive interactive systems are radically different than standard GUIs, largely because of the nature of human communication, and their basic architecture reflects these differences [8]. Pen-based adaptive interfaces are also expected to be easier to learn and use to improve user's satisfaction. They have the potential to expand computing to more challenging applications, to be used by a broader spectrum of everyday people, and to accommodate more adverse usage conditions than in the past. Pen gestures and adaptive interface can be efficient in the visual-spatial application, such like the geometry application or the intelligent whiteboard, involving many kinds of domain. We plan to do deep research on the dynamic human and human interaction in the class based on pen gesture and adaptive users' intent, further to reduce the workload of communication between teacher and students in the class.

There are many difficulties remained. In some conditions, pen can take the place of mouse, but what fits in with a mouse is not similar with a pen, and vice versa. For any given task, the ideal pen-based UI should not be limited to technique developed for GUIs, but incorporate pen-specific technique that takes advantage of the special characteristic of pen. Gestures invoke operations for direct manipulation to objects. At the same time, it is difficult to recognize users' intent well enough for application. Whereas misrecognition of a character is easily detected by users, misrecognition of an operator may not be. Furthermore, an unintended operation is likely to be more difficult to be corrected than an incorrectly recognized character [9].

Acknowledgments. This work was supported by the National Key Basic Research and Development Program under Grant No.2006CB303105 and by the National Science and Technology Supporting Program under Grant No. 2006BAF01A44.

References

1. Gregory, D.A., Elizabeth, D.M.: Charting past, present and future research in ubiquitous computing. ACM Transactions on Computer-Human Interaction 7, 29–58 (2000)
2. Oviatt, S.L., Cohen, P.R., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J., Ferro, D.: Designing the user interface for multimodal speech and gesture applications: State-of-the-Art systems and research directions. Human Computer Interaction 15, 263–322 (2000)
3. Francisco, B., José, L.V.: A Dynamic-symbolic interface for geometric theorem discovery. Computers & Education 38, 21–35 (2002)
4. Dey, A.K, Abowd, G.D., Salber, D.: A Context-based infrastructure for smart environments. In: The Proceedings of the 1st International Workshop on Managing Interactions in Smart Environments, pp. 114–128 (1999)

5. Long, A.C., Landay, J.A., Rowe, L.A.: PDA and gesture use in practice: Insights for designers of pen-based user interfaces. Tech. Rep. UCB//CSD-97-976, U.C. Berkeley (1997)
6. Allan Jr.C.L.: Improving gestures and interaction technology for pen-based user interface. The. Proceedings of CHI98 18–23 (1998)
7. Oviatt, S.L., DeAngeli, A., Kuhn, K.: Integration and synchronization of input modes during multimodal human-computer interaction. In: Proceedings of the CHI'97, pp. 415–422. ACM Press, NY (1997)
8. Ma, C.X., Dai, G.Z., Teng, D.X., Chen, Y.D.: Gesture-based interaction computing in conceptual design. Journal of Software 16, 303–308 (2005)
9. Ma, C.X., Dai, G.Z., Wang, H.A.: Perceiving the invisible: Natural interactive collaboration based on sketch in conceptual design. In: The Proceedings of the 7th Asia-Pacific Conference on Computer-Human Interaction. Taipei, Taiwan (2006)
10. Richard, A., Ruth, A., Beth, S., Steven, A., Ken, Y.: Experiences with a Tablet PC based lecture presentation system in computer science courses. In: The Proceedings of the 35th SIGCSE Technical Symposium on Computer Science Education, Norfolk, Virginia, USA (2004)
11. Evan, G.: Handwritten slides on a Tablet PC in a discrete mathematics course. In: The Proceedings of the 35th SIGCSE Technical Symposium on Computer Science Education, ACM SIGCSE Bulletin 36 (2004)
12. Michelle, W., William, G.G., Beth, S.: Ubiquitous presenter: Increasing student access and control in a digital lecturing environment. In: The Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education, ACM SIGCSE Bulletin 37 (2005)

Part II

Cognitive Workload and Human Performance

Asymmetric Synchronous Collaboration Within Distributed Teams

M. Ashdown and M.L. Cummings

Humans and Automation Laboratory, Massachusetts Institute of Technology,
77 Massachusetts Avenue, Cambridge, MA 02139, USA
http://web.mit.edu/aeroastro/www/labs/halab/

Abstract. Teams performing physical tasks must often be distributed in space, and are often organized hierarchically. This means that systems to support collaboration between members must account for the asymmetry in physical environment, organizational roles, and available technology. Using urban search and rescue as an example, we first describe the factors that cause this asymmetry. We then discuss the way information should be shared, and the type of awareness that should be supported. We suggest the use of very different display and interaction devices for operators at the organizational levels, to complement their situations and needs.

Keywords: synchronous remote collaboration, computer-supported cooperative work, command and control.

1 Introduction

A team of people with a task is often organized in a hierarchy, to allow it to scale with the magnitude of the task. This is especially true for command and control organizations in the military, and emergency response in the civilian sector. Team members must often be distributed geographically, so they communicate via radios, and more recently, with digital communications.

In this paper we consider situations in which a real-world task is performed, our example being urban search and rescue (USAR). The goal will be to search damaged buildings and extract victims. The search and rescue will be conducted by a combination of humans and robots with operators in the field, and these will be linked via a digital network to personnel at a base who supervise and plan the operation (Figure 1). The physical domain is dynamic so the personnel must react to it under time-pressure. We assume a network is available and that the team has access to technology that will be available in the near future. We address the problem of how technology should be designed to support synchronous collaboration between personnel in the field and those at the base.

Personnel in the field will be in a different environment, using different equipment, and performing different tasks from those at the base. This means that

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 245–255, 2007.

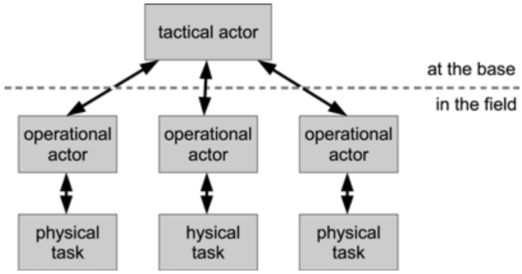


Fig. 1. We consider situations where a tactical actor is situated in a base and is supervising operational actors in the field. In the case of USAR, an operational actor's physical task may be searching and rescuing directly, or controlling a robot.

the information required by the the two classes of collaborators will be very different, as will the technology that can best support them.

In this paper, we will describe the nature of the asymmetry inherent in hierarchies used to manage personnel in the field, highlight some issues for design of remote collaboration for this type of task, and suggest the combination of different display types and interfaces for creating a distributed synchronous collaboration system that supports the information needs and awareness of different classes of personnel.

2 Hierarchical Organization

Levels in a command hierarchy are typically labelled *operational*, *tactical*, and *strategic*: operational units perform their duties in the field, the tactical level allows multiple operational units to be supervised, and the strategic level makes long-term decisions governing the mission.

We are considering hierarchical organizations with distributed personnel that collaborate synchronously using voice and data channels, to operate in a dynamic domain. Examples include military command and control organizations, and the Incident Command System [3] used in crisis response. Other relevant domains include air traffic controllers and pilots, where Data Link communications have reduced time spent and distance flown [26], and dispatch of vehicles such as taxis, which have benefited from the use of GPS and digital networking [16].

The agents in such a system have conventionally been human actors, but these are now joined at the operational level by uninhabited vehicles (UVs), including robots, which must often be accompanied in the field by their operators. Conventionally, communication between locations has been achieved by speaking on radios, but synchronous digital exchange is now becoming available. Even in domains like crisis response where infrastructure may have been damaged, mesh networks can be formed by combining multiple portable devices and mobile network nodes positioned around an incident site. Sensing systems such

as the global positioning system (GPS) can provide real-time data about the current situation.

Operational personnel will be working under time pressure, possibly in noisy, hazardous environments with limited ability to carry or use large, complex computing devices. In USAR, the primary task will be to search and rescue, or operate robots to search for victims using video cameras and other sensors. Communicating with the tactical level is an operator's secondary task, so it should not impede the primary one, and the communication device should be compact, unobtrusive, and simple to operate. Tactical personnel must deal with large amounts of information to form awareness of the whole situation, and then filter and fuse it before communicating it to the operational personnel. Thus, support should be provided for tactical actors to view a large amount of information simultaneously, but coherently.

Most remote synchronous collaboration technologies, such as videoconferencing and groupware [10], assume that the situations and roles of the collaborators are fairly symmetric, and that the data they share is not subject to external influences. In contrast, as discussed below, we are considering a class of domains where significant asymmetry exists between collaborators, and where the situation is dynamic: it progresses independently of the collaborators and it causes time-pressure. Advances in displays, networks, and sensing will make remote collaboration involving a multitude of real-time data possible in these domains. The design of collaboration systems should be driven by the fundamental physical, organizational, and technological asymmetries.

3 Role Asymmetry

There are major differences between the situations of the operational and tactical actors, and even between operational actors, performing a manual task is different from controlling a UV (Table 1). UV control falls between the other two conditions in many respects.

Role. The operational actor is situated in the field where the physical task occurs. This may be performed directly, or in our USAR example may involve

Table 1. Differences between operational and tactical situations

	operational		tactical
	manual	UV	
role	execute actual task	control robot to do task	supervise operations
environment	difficult, hazardous	noisy, distracting	controlled
latency	< 1 second	seconds	minutes
knowledge of world	local, detailed	local, depends on sensors	global, coarse
knowledge of plans	narrow	narrow, but plans available	comprehensive
computing	portable, limited	portable	fixed, powerful

the use of a robot. In contrast, the tactical actor's job is to supervise: she deals with monitoring, planning, and commanding.

Environment. The environment at the operational level presents various constraints because it is determined by external factors, may be restrictive in various ways, and may well require the operational actor to move from location to location. The environment in the base of operations for the tactical actor can be chosen to fit the needs of the task: we will assume it is fixed for the duration of the mission, is large enough to contain all the desired technology, and is protected from weather and other hazards.

Latency. Because the operational actor interacts directly with the external world, feedback will be fast, and the classical observe-orient-decide-act loop [9] will occur rapidly. For UV control, communication latency will affect this loop, thus degrading the human's interaction with the world. The tactical actor will work on longer timescales, and it may take some time to receive feedback on the effectiveness of a plan being executed. It may be possible for the tactical actor to take control of a UV from afar, but the end-to-end latency and reliabilty of the network will limit this.

Knowledge of world. An operational actor will possess detailed knowledge of the physical situation around him. The tactical actor will form a wider view that does not contain such fine detail, but does contain significantly more information from a wider variety of sources. A UV operator's view will depend on the sensors available: for instance, proprioceptive sensors that sense which direction is up, and whether a robot's wheels are turning, will help to diagnose problems. Information must be aggregated as it is passed up from operational to tactical to avoid information overload, and similarly the tactical actor needs to select which information is useful before contacting any particular operational actor.

Knowledge of plans. The tactical actor is charged with forming and maintaining a plan for the whole team, so she will have knowledge of the whole plan. An individual operational actor needs to know the details, only for the part corresponding to her own task. The tactical level should transmit updates to the plan to the operational level as needed. Compared to a manual operational actor, because a UV operator's primary task is performed with a computer, she should have more opportunity to stay aware of changes to the plan displayed on her screen.

Computing. The physical constraints at the operational level mean that any computing device must be small and portable, so the power of the device and the amount of information it can display will be limited compared to the tactical level. The tactical actor may be in a permanent command centre, or a mobile base of operations, but in either case state of the art computing and display facilities should be available.

Velez et al. [27] found that in remote collaboration, role asymmetry combined with platform heterogeneity impacts collaboration. For instance, they found that in a two-person system, when one collaborator has a restricted view of a task

due to hardware constraints, some status is conferred onto the other, who will have a tendency to take charge. Technological constraints, such as asymmetric processing power and network bandwidth, will affect the design of collaboration systems [13]. However, other fundamental constraints should be addressed, such as differences in attention available from the user, differences in display devices, and noisy environments at the operational level.

4 Remote Collaboration

We are considering situations where the tactical actor collaborates remotely with the operational actors he is supervising. Advances in networking will allow the medium for collaboration to consist of data and graphics, in addition to the conventional voice channel. Below we describe some issues in designing additional collaboration channels, and we emphasize the importance of supporting awareness, both at the operational and the tactical level.

The case of UV tasking provides an extra possibility over conventional command hierarchies, because tactical personnel may be given the ability to take control of some or all of the capabilities of the UVs at the operational level, subject to network limitations. Also, operations involving heterogeneous systems may involve the tactical operator supervising some UVs directly while collaborating with operational personnel controlling other UVs in the field.

4.1 Sharing Information

The standard audio channel used in many command and control situations may now be augmented with a digital channel. Using cameras for collaboration will not be very useful in this case; instead, a digital channel can be used to share visual information that represents task status, acting, and planning. This can be called a task space, as opposed to a person space [4]. It allows the collaborators to transmit information more precisely than when speaking, ground their conversations on the information they share, and perform visual gestures that may be much simpler than their verbal equivalents.

A geographic information system (GIS) displaying maps and GPS locations is a prime example of the type of large-scale dynamic data that can be displayed at the tactical level to aid in situation assessment and planning. In our USAR example, a dynamic map might contain buildings to be searched, the locations of operational personnel and robots, and additional information useful for the planning activities. The tactical actor will want to view the whole map. During synchronous collaboration with an operational actor, a portion of the map may be shared between the two to allow deictic references; collaborators can point and annotate as they talk over the voice channel about plans and actions.

Mulgand et al. [17] consider data shared at three distinct levels: *pixels*, *data*, or *recipes*. Sharing pixels creates a *what you see is what I see* system, provides

common ground in the conversation, but restricts the collaborators to having identical views. Sharing data allows collaborators to see customized views of the same data. Sharing recipes is the most flexible because each collaborator can modify the recipe—this is essentially the database query used to generate the data. However, this means collaborators may end up viewing quite different representations of the same situation, so they will have to rely more heavily on the voice channel to resolve discrepancies.

4.2 Awareness

It is important to support awareness in a collaborative system. A dynamic domain will require awareness of both the changing situation in the external world, and of the data and plans that the collaborators share. Gutwin and Greenberg [11] define four basic characteristics of awareness:

1. Awareness is knowledge about the state of a particular environment.
2. Environments change over time, so awareness must be kept up to date.
3. People maintain their awareness by interacting with the environment.
4. Awareness is usually a secondary goal—that is, the overall goal is not simply to maintain awareness but to complete some task in the environment.

Workspace awareness [11] means knowing what other collaborators are doing in a shared workspace, and can be supported by views of the actions of others such as radar views, which show a miniature version of a whole workspace so that the actions of others are visible. The deictic references mentioned in the section above are a form of intentional communication, whereby a collaborator explicitly signals to another, but consequential communication, whereby one collaborator becomes aware of the actions of another by watching them happening, is also important. This happens automatically during collocated collaboration, but in a distributed system, features to support this type of communication must be explicitly added. One example is *feedthrough*: for example, a graphical object can give visual feedback of the operations performed on it, for the benefit of remote observers. This could be used if one collaborator was adding points of interest to a map, to allow another to watch them being created. In addition, network reliability and latency will effect the design of the shared task space. For example, higher latency will require more persistent remote gestures because it will be harder for collaborators to synchronize with each other.

In a dynamic domain it is not just workspace awareness regarding shared digital data that is necessary, but also *situation awareness* [7]: knowing what is going on in the external world. This consists of perceiving the elements of the world, comprehending what they mean, and predicting future states. Generally it is a challenge to get information passed up from the operational to the tactical level, to allow perception of the situation at the higher level. Advances in portable computational devices and sensing technologies will allow much data on the state of operations to be automatically transmitted in real-time up to the tactical level. However, because the tactical actor is working at a higher level of abstraction, it will probably be necessary to distil the low-level data into a status

visualization that can be rapidly assessed. In particular, the many sensors fitted to UVs will allow large amounts of information to be sent up to the tactical level, but visualizations should be devised that avoid cognitively overloading tactical personnel.

It is important to note that collaborators must remain aware, not only of raw data and the external situation, but also of the plans of the agents in the system, and how well they are fulfilling them. This could be called *activity awareness* [5], and presumes a method for actors to express their plans to the computer.

5 Technology

Ubiquitous computing as envisaged by Weiser [29] occurs in an environment populated by computing devices of different sizes—*tabs*, *pads*, and *boards*. Standard computer screens are pad-sized, at 15-21 inches. The tab-sized, 2-4 inch displays of mobile phones and personal digital assistants have become popular, and recently there has been significant research interest in the board-sized displays, of 60-inches or more, that can now be constructed. The quantitative size differences between these devices entail qualitative differences in the activities they may support.

We propose supporting asymmetric collaboration between actors in a distributed, organizational hierarchy by combining very different computing devices. At the operational level, personnel will use use small displays on portable devices, due to the constraints of the physical situation and because their organizational role requires that they spend most of their time on their physical task rather than manipulating data. At the tactical level, personnel will use large displays that allow a significant amount of data to be monitored and manipulated. The data will consist of information about the current situation, such as maps of a search site with locations of the units, and future plans, such as lists of buildings to search and victims to extract and treat.

5.1 Portable Devices

Advances in areas such as batteries, wireless networks, and small screen technology are fuelling an upsurge in the use of portable computing devices, but Raghunath et al. [18] argue that "advances in technology will not significantly mitigate handhelds' limitations because human perceptual and motor systems—not the underlying technology—are the real limiting factors." These factors include human visual acuity, which limits the amount of information that can be displayed on a handheld, and size, which makes it impractical for multiple people to view or interact with such a small device.

The use of handheld computers in the field that allow real-time communication between remote actors has been pioneered by pervasive games [2]. Handhelds have also been used successfully in medical settings [25]. There has been some

work on combining patients' screens, which are primarily used for entertainment, with such handhelds [1].

5.2 Large Displays

In the controlled environment of a base of operations, tactical actors will be free to exploit the benefits of the large display devices that are becoming available. Rather than the low-resolution shared wall displays that are often used in command centres, we mean large high-resolution personal displays. The Office of the Future [19] pioneered the use of large displays in an office environment. Infrastructure for large displays has received considerable attention [8], and includes multi-projector calibration, and systems for visualization of large datasets. LambaVision [15] is an architecture for combining an array of flat-panel monitors into a single large display; currently the seams between monitors are visible, but advances in hardware may soon allow them to disappear. Interfaces for such displays have recently received much attention, particularly for collocated collaboration [21], and recent developments in multi-touch sensing [12] hold the promise of more fluid interfaces.

A large display is qualitatively different from a small one, and large displays have been shown to have benefits for cognition [23] and productivity [6]. Swaminathan & Sato [22] cite the ability to collaborate with several other people, and the ability to present a primary work object in context as a important benefits of such displays. They state that a large contiguous display, such as an interactive desk, is most useful for applications with the following characteristics:

1. A large amount of interrelated information needs to be displayed, and
2. any part of this information could, at any time, become the centre of user's attention and may need to be carefully studied or modified.

In our USAR task, such a device would allow the simultaneous display of detailed maps, plans for the operation, and status of operational units. All this information must be considered at the tactical level, but can be condensed into a simple form for each operational unit, such as a prioritized list of buildings to search.

Large displays have different affordances as compared to conventional ones, and pose different challenges. For instance, a mouse pointer can easily get lost, window management requires different mechanisms, and conventional menus are awkward to use when the menu bar is far away. A large display, even if technically homogenous, will not be used in a homogenous manner by the user, and the much larger space makes it possible to simultaneously display much more information, which enables more complex multitasking behaviour [20]. The realization that the standard Window Icon Menu Pointer (WIMP) interface does not scale to very large displays has led to a series of modifications to the standard interface, and also to some radical re-inventions, including multi-touch finger and pen input, gesture recognition, arbitrarily rotated and scaled items placed in 'bins', user interface widgets that respond differently to each user, and physical objects used to control the digital content [14,21].

5.3 Collaboration Between Screen Sizes

Screen size affects performance in tasks, for instance the lack of context when looking up items in a large table on a small screen reduces performance [28]. Tremaine et al. [24] had remote users collaborate synchronously on a puzzle task using either a conventional personal computer or a handheld computer, with the handheld offering a more restricted view of the problem. The results in the PC-handheld and PC-PC conditions were better than those in the handheld-handheld condition, because if at least one person had a good view of the task by using the PC, they could help the other person. In that case the person on the PC had the necessary situation awareness, which was remotely shared.

When using a shared view of a map, the tactical actor with the large display may be given control of the viewing parameters because she has the context necessary to make informed decisions about where to look. In general, the tactical actor may naturally assume much of the control of the shared data because this is the domain in which she works, while the operational actor is primarily dealing with his physical task. Thus the use of different display devices will complement the physical environments and organizational roles of the collaborators.

6 Conclusion

In a hierarchical crisis management organization, it is desirable to have tactical actors situated in a secure base, but operational actors must often be located in the field. Advances in displays, networks, and sensing will allow synchronous remote collaboration between these two levels. However, organizational asymmetry, and cognitive and physical constraints at the operational level, will mean that technology for collaboration should reflect the differing roles and situations of the collaborators.

A task space between distributed collaborators can be used to share information that is difficult to convey verbally. The level of sharing—pixels, data, or recipes—must be chosen. Also, support should be provided for the workspace, situation, and activity awareness of all actors in the system. We have proposed the use of display devices of very different sizes at the different organizational levels. At the operational level, small, simple, portable devices will support the viewing of individual plans and indication of status information. At the tactical level, a large display will allow for the comprehension of the complete situation and formulation of a plan that includes multiple operational units.

In our future work we will implement a USAR system for synchronous remote collaboration using different display devices to complement the work of actors in a hierarchical organization, as described above. We will design interfaces to support the types of awareness listed above, and perform experiments to validate our design and provide recommendations for similar systems.

Acknowledgements. We thank Gilles Coppin and Stacey Scott for their valuable comments and insights. This research was conducted in conjunction with

Thales Research and Technology (UK) and funded by European Commission FP6 Marie Curie Outgoing International Fellowship number 21743.

References

1. Alsos, O.A., Svanaes, D.: Interaction Techniques for Using Handhelds and PCs Together in a Clinical Setting. In: Proc. NordCHI, pp. 125–134 (2006)
2. Benford, S., Magerkurth, C., Ljungstrand, P.: Bridging the Physical and Digital in Pervasive Gaming. Communications of the ACM 48(3), 54–57 (2005)
3. Bigley, G.A., Roberts, K.H.: The Incident Command System: High-Reliability Organizing for Complex and Volatile Task Environments. Academy of Management Journal 44(6), 1281–1299 (2001)
4. Buxton, W.: Telepresence: Integrating Shared Task and Person Spaces. In: Proceedings of Graphics Interface '92, pp. 123–129 (1992)
5. Carroll, J.M., Neale, D.C., Isenhour, P.L., Rosson, M.B., McCrickard, D.S.: Notification and awareness: synchronizing task-oriented collaborative activity. Int. J. Human-Computer Studies 58(5), 605–632 (2003)
6. Czerwinski, M., Smith, G., Regan, T., Meyers, B., Robertson, G., Starkweather, G.: Toward Characterizing the Productivity Benefits of Very Large Displays. In: Proceedings of IFIP Interact 2003, pp. 9–16 (2003)
7. Endsley, M.R.: Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors, 37(1) (1995)
8. Funkhouser, T., Li, K.: Large-Format Displays (introduction to special edition on large displays) In: IEEE Comp. Graph. & App. pp. 20–21 (2000)
9. Grant, T., Kooter, B.: Comparing OODA & Other Models as Operational View C2 Architecture. In: Proc. 10th ICCRTS (2005)
10. Grudin, J., Poltrock, S.E.: Computer-Supported Cooperative Work and Groupware. Advances in Computers 45, 269–320 (1997)
11. Gutwin, C., Greenberg, S.: Salas, E., Fiore, S.M., Cannon-Bowers, J.A. (eds.) Team Cognition: Process and Performance at the Inter- and Intra-individual Level, chapter The Importance of Awareness for Team Cognition in Distributed Collaboration. APA Press (2004)
12. Han, J.Y.: Low-Cost Multi-Touch Sensing through Frustrated Total Internal Reflection. In: Proceedings of UIST 2005, pp. 115–118 (2005)
13. Krebs, A.M., Dorohonceanu, B., Marsic, I.: Supporting Collaboration in Heterogeneous Environments. J. Management Inf. Sys. 20(4), 199–227 (2004)
14. Kurtenbach, G., Fitzmaurice, G.: Applications of Large Displays (introduction to special edition). IEEE Comp. Graph. & App. 25(4), 22–23 (2005)
15. Leigh, J., et al.: The Global Lambda Visualization Facility: An International Ultra-High-Definition Wide-Area Visualization Collaboratory. In: Future Generation Computer Systems, 964–971 (2006)
16. Liao, Z.: Taxi Dispatch vis Global Positioning Systems. IEEE Transactions on Engineering Management 48(3), 342–347 (2001)
17. Mulgand, S., Travis, A., Standard, J., Means, C., Burgman, A.: Shared User Interfaces for Dynamic Collaboration in Network-Centric Command Centers. In: Proc. ICCRTS (2005)
18. Raghunath, M., Narayanaswami, C., Pinhanez, C.: Fostering a Symbiotic Handheld Environment. IEEE Computer, 56–65 (2003)

19. Raskar, R., Welch, G., Cutts, M., Lake, M., Stesin, L., Fuchs, H.: The Office of the Future: A Unified Approach to Image-Based Modelling and Spatially Immersive Displays. In: Proceedings of ACM SIGGRAPH 98, pp. 179–188 (1998)
20. Robertson, G., Czerwinski, M., Baudisch, P., Myers, B., Robbins, D., Smith, G., Tan, D.: The Large-Display Experience. IEEE Comp. Graph. & App. 25(4), 44–51 (2005)
21. Scott, S., Carpendale, S.: Interacting with Digital Tabletops (introduction to special edition). IEEE Comp. Graph. & App. 26(5), 24–27 (2006)
22. Swaminathan, K., Sato, S.: Interaction Design for Large Displays. Interactions 4(1), 15–24 (1997)
23. Tan, D.S., Gergle, D., Scupelli, P.G., Pausch, R.: With Similar Visual Angles, Large Displays Improve Spatial Performance. In: Proc. of CHI 2003, pp. 217–224 (2003)
24. Tremaine, M., Sarcevic, A., Wu, D., Velez, M.C., Dorohonceanu, B., Krebs, A., Marsic, I.: Size Does Matter in Computer Collaboration: Heterogeneous Platform Effects on Human-Human Interaction. In: Proc. 38th Annual Hawaii Int. Conf. System Sciences (HICSS 2005), p. 33 (2005)
25. Turner, P., Milne, G., Kubitscheck, M., Penman, I., Turner, S.: Implementing a Wireless Network of PDAs in a Hospital Setting. Per. Ubiq. Comp. 9(4), 209–217 (2005)
26. User Benefits of Two-Way Data Link ATC Communications. Technical Report DOT/FAA/CT-95/4, US Department of Transportation (1995)
27. Velez, M., Tremaine, M., Sarcevic, A., Dorohonceanu, B., Krebs, A., Marsic, I.: Who's in Charge Here? Communicating Across Unequal Computer Platforms. ACM Trans. CHI 11(4), 407–444 (2004)
28. Watters, C., Duffy, J., Duffy, K.: Using Large Tables on Small Display Devices. Int. J. Human-Computer Studies 58(1), 21–37 (2003)
29. Weiser, M.: The Computer for the 21st Century. Scientific American, 94–104 (September 1991)

Situation Awareness and Secondary Task Performance While Driving

Martin R.K. Baumann, Diana Rösler, and Josef F. Krems

Chemnitz University of Technology, Institute of Psychology,
D-09107 Chemnitz, Germany
{Krems,Diana.Roesler,Martin.Baumann}@phil.tu-chemnitz.de

Abstract. For safe driving it is necessary that the drivers perceive the relevant objects of a situation, comprehend the meaning of these objects to form a holistic understanding of the current situation, and predict the future development of the situation. A concept that aims to describe and integrate these different cognitive processes is situation awareness, for example [1]. According to this concept it is assumed that a mental representation is constructed, maintained, and updated while driving. Attentional and working memory (WM) resources are involved in these processes. If secondary tasks performed while driving impose significant load on visual attention and WM, then one can expect that situation awareness is impaired. We investigated these predictions in 2 experiments. The results show that both cognitively and visually demanding tasks interfere with the maintenance of a correct situation model in memory. Visually demanding tasks do not always seem to interfere with visual attention processes in ways that lead to degraded situation awareness.

Keywords: Visual demand, situation awareness, secondary tasks, driving.

1 Introduction

For safe driving it is necessary that drivers perceive, identify, and correctly interpret the relevant objects and elements of the current traffic situation and that they construct expectations about the future development of the current situation to adapt their own driving behavior to the situation. Such elements may be other traffic participants, the surface of the street, or traffic signs. *Situation awareness* has been recently proposed in aviation psychology as a concept that aims to describe and integrate these different cognitive processes within a common framework ([1], [2], [3]).

According to Endsley three functions of situation awareness can be distinguished [1]. The first function involves the perception of the status, the attributes, and the dynamics of the relevant situation elements. The second function, comprehension, aims to integrate the different situation elements into a holistic picture of the situation, resulting in the comprehension of the meaning of the different elements. The third function of situation awareness is to construct expectations about the future behavior of the situation elements on the basis of the comprehension of the situation.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 256–263, 2007.

Whereas Endsley [1] describes different cognitive resources and mechanisms that might be involved in constructing and maintaining situation awareness her model is rather vague about the nature of these processes and how the different functions of situation awareness are realized. Baumann and Krems [4] proposed on the basis of Kintsch's theory of discourse comprehension [5] a model of situation awareness that assumes that situation awareness is constructed by a comprehension process (cf. [2]). Perceived information activates knowledge stored in long-term memory (LTM) that is linked to the perceived information. From this activated knowledge network a coherent representation is constructed by a constraint satisfaction process. This process constrains the spreading of activation leading to the activation of compatible and to the suppression of incompatible knowledge elements of the activated knowledge network. The result is a coherent mental representation of the current situation, the situation model. For example, an event such as a traffic light turning yellow might activate at first two incompatible interpretations, “I have to decelerate to stop before the traffic light” and “I have to accelerate to pass the crossroads before the traffic light turns red”. These two interpretations might receive additional activation from other knowledge elements. For example, if the driver knows that the police monitor this crossroads, that knowledge will additionally activate the deceleration interpretation and at the same time inhibit the acceleration interpretation. With this additional activation the deceleration interpretation “wins” and the acceleration interpretation will be inhibited. The experienced driver's knowledge presumably also includes expectations about the future development of situations that are linked to certain types of situation and activated when these types of situations, such as approaching a traffic light, are encountered. In this sense the same process that serves the comprehension function of situation awareness also serves the prediction function, especially in routine driving situations for which the driver already acquired relevant knowledge. The situation model is the basis for the driver's actions. These actions and the dynamics of the driving task make it necessary to update the situation model constantly.

This comprehension based view of situation awareness emphasizes the importance of cognitive resources for the construction and maintenance of situation awareness besides visual attention. The integration of new information into the situation model, the updating of the model, and the use of the situation model as a basis for action selection all need working memory (WM) resources. Visual perception and attention processes serve the perception function of situation awareness. Both the driver's visual and cognitive resources need to be available to construct a proper situation model that is necessary for driver's safety. But in modern cars more and more driver support and information systems are available to the driver confronting the driver with more and more in-car tasks. Performing these in-car tasks while driving creates a dual task situation for the driver that could lead to the impairment of situation awareness if the task strongly involves visual attention or cognitive resources. An in-car task that is highly visually demanding may lead to a degraded situation model as many relevant elements of the traffic situation will simply not be perceived and therefore not integrated into the situation model. A highly cognitively demanding in-car task should interfere with the activation of knowledge, the maintenance of the situation model in WM, and the updating of the model. In the design process of such in-car tasks it is therefore important to consider the effects of such tasks on situation

awareness to avoid too visually and / or cognitively demanding tasks. The first goal of this paper is to describe a procedure that should allow to evaluate the visual and cognitive demands associated with such tasks. A sample of tasks was evaluated with this procedure. Then this sample of tasks had to be performed while driving and the effects of these tasks in driver's situation awareness was assessed. The second goal of this paper is accordingly to test the predictions of the comprehension based situation awareness model [4] about the effects of visually and cognitively demanding tasks on situation awareness.

The evaluation procedure uses a dual-task technique to measure the visual and cognitive demands of in-car tasks. The basic idea is to measure the degree of interference between the processes involved in the construction and maintenance of situation awareness and those that are involved in performing in-car tasks. Participants have to perform the in-car tasks in the laboratory as primary tasks. The secondary task is used to measure both the residual visual attention and WM resources not utilized by the primary in-car task. We used a 1-back task as secondary task. In this task participants are presented with one of four different visual stimuli on each trial. The participants' task is to indicate whether the current stimulus is identical to the one presented in the previous trial. This task seemed suitable for assessing situation awareness relevant interference effects of in-car tasks for the following reasons: The reaction to a stimulus is context dependent, that is a minimum situation model has to be kept in memory and this situation model has to be updated frequently. Thus, the task involves two basic components of situation awareness processes.

The 1-back task yields two measures that are supposed to allow assessment of the visual and cognitive demands of the primary task. First, one can look at the proportion of visual stimuli responded to, not differentiating whether the response was correct or not. This *detection rate* is used as a measure of the visual demand of the in-car task. If the in-car task is highly visually demanding the participant will simply miss many of the visual stimuli. In this regard, the task is similar to a laboratory variant of the peripheral detection task (PDT, [6]) that was previously developed to measure only visual demand of in-car tasks. The proportion of correct responses of *all* shown responses, *the hit rate*, is used as a measure of the cognitive demands of the in-car task. If the task is highly demanding the participant may often forget the previous stimulus or may frequently fail to update the stimulus. This should lead to more false responses to the visual stimuli.

2 Experimental Studies

We conducted two experiments. In the first experiment three tasks differing in visual and cognitive demand were evaluated with the procedure described above. In the second experiment these tasks were performed while driving in a driving simulator and different measures of situation awareness were collected. The aim of these two experiments was to test the prediction that tasks that were evaluated as highly visually demanding should show an impairment of the perception function of situation awareness whereas tasks that were evaluated as highly cognitive demanding should show an impairment in maintaining and updating the situation model.

2.1 Experiment 1

The aim of Experiment 1 was to evaluate the visual and cognitive demands of different tasks using a 1-back task as the measurement procedure, where the participants had to compare the current stimulus with the one that was presented immediately before.

Method. In this experiment 21 participants performed three in-car tasks together with a 1-back task. The participants' age ranged from 19 to 30 years. There were 11 women and 10 men.

The in-car tasks were *Listening* to a spoken text, asking an automated train information system with automated *Speech* recognition for the arrival time of a train, and writing a short text with a *PDA* using its touch screen and stylus.

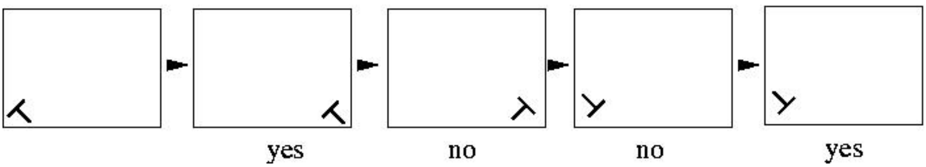


Fig. 1. Example of a 1-back task stimulus sequence with the correct answer for each stimulus presentation

In the 1-back task the participants were presented with T-like visual stimuli that were rotated by 45°, 135°, 225°, or 315° (see Fig. 1). The size of the stimuli was 1.2° of visual angle. Each stimulus was presented for 500 msec. The interval between two stimulus presentations varied randomly between 1 and 2 sec. They were presented at either 23° of visual angle to the left or to the right of the participant at a distance of 2.5 m. The participants' task was to decide for each new stimulus whether it was identical to the one presented in the trial before. The probability for each new stimulus being identical to the one presented before was .5. In this task each stimulus acts as the context for the next stimulus and the context could change after each stimulus presentation. Therefore, for each stimulus presentation participants had to decide whether this stimulus was identical to the previous one and then they had to encode the new stimulus as the new context, get rid of the old context, keep the current context in WM until the next stimulus was presented, and then encode this stimulus as the new context. As explained above the detection rate was used as a measure of the visual demand of the respective in-car task and the hit rate was used as a measure of the cognitive demand of the in-car task.

Table 1. Detection and hit rate in the 1-back task for the different in-car tasks

	Detection rate	Hit rate
Baseline	.84	.89
Listening	.78	.84
Speech interaction	.68	.77
PDA writing	.17	.47

Results and Discussion. The detection and the hit rates for the different in-car tasks are shown in Table 1. The detection rates of the different tasks differed significantly, $F(2.25, 42.76) = 121.67$, $p < 0.001$. All tasks were significantly different from each other, with one exception: baseline and listening did not differ significantly. As expected the task that involved the interaction with a visual display, the PDA writing task, had the lowest detection rate. Additionally, the difference between the listening and the speech interaction task was also significant. As there is no difference in the visual demand of both tasks this difference can only be attributed to the difference in cognitive demand between tasks. Therefore, this difference indicates that the detection rate is to some extent also sensitive to the cognitive demands of tasks.

Only the hit rate data from 17 participants were analyzed as four participants showed no reaction to the 1-back task stimuli during the PDA writing task. As for the detection rate the tasks showed an overall difference, $F(1.37, 21.93) = 21.23$, $p < 0.001$. There was no significant difference between baseline and listening. All other tasks had a significantly lower hit rate than the baseline. Additionally, the listening and the speech task significantly differed from the PDA writing task.

The results indicate that detection rate might be a valid measure of the visual demand of tasks, although it is also influenced by the cognitive demand of the tasks. The interpretation of the hit rate as a measure of the cognitive demand of tasks is much more difficult. First, the task that was associated with a low detection rate (PDA writing) had also a low hit rate. Second, there was no significant difference between the listening and the speech interaction task despite the fact that the cognitive demands of both tasks clearly differ and that there was a significant difference between both tasks in the detection rate that was attributed to the different cognitive demands of both tasks. It might be the case that the hit rate of the 1-back task was not sensitive enough to detect the difference in cognitive demand between the listening and the speech task.

2.2 Experiment 2

Experiment 2 was a driving simulator study. The aims of this experiment were to examine the effects of the in-car tasks of Experiment 1 on situation awareness and to compare these effects with the effects these tasks had on the detection and the hit rate of the 1-back task in Experiment 1. The driving scenario consisted of driving in the middle lane of a motorway with three lanes in each direction. The participants were instructed to drive in the middle lane at 110 km/h whenever possible. Situation awareness was measured by repeatedly interrupting the driving simulation and asking the driver questions about the number of cars in different lanes. There were four relevant locations that were tested. The four locations were: i) left lane behind the driver, ii) left lane in front of the driver, iii) middle lane in front of the driver, and iv) right lane in front of the driver. After each interruption only one of the four possible locations was tested.

The response accuracy for the locations in front of the driver should indicate how the perception function of situation awareness is influenced by an in-car task. This information is always visible and incorrect answers should mainly arise from failures in perception and visual attention. The response accuracy for the location behind the driver should indicate the effects of the in-car tasks on the maintenance and updating

of the situation model in WM. This information has to be kept in memory, presumably in working memory, until the driver updates the information by looking into the interior or the left external mirror. Incorrect answers should arise from the driver either not looking into the mirrors or forgetting what was encoded before. Therefore, we expected first that the visually demanding PDA writing task should lead to more incorrect answers to questions regarding the front locations than the not visually demanding tasks listening and speech interaction. Second, we expected that both visually and cognitively demanding tasks, both the PDA writing and the speech interaction task, should lead to more incorrect answers to questions regarding the rear location than the less cognitively demanding task listening.

Method. 19 participants performed the same three in-car tasks that were used in Experiment 1 while driving on a motorway in a driving simulator. 15 of the participants were female. The mean age was 23.79 (SD = 2.82).

For the driving simulation we used the driving simulator of Systems Technologies Inc. The simulation was presented on three 19”-TFT screens providing a projection of 135° of visual angle.

Participants performed twelve trials of each in-car task. For each trial a new ride was started. During each ride some critical events happened, such as braking of the lead car, to make sure that the participants paid attention to the driving task. After 2000 m, 3000 m, or 4000 m of driving the simulation was suddenly interrupted and the participants were asked about the number of cars on one of the four positions mentioned above. Each position was tested three times during each in-car task block.

Results and Discussion. A 4 (locations) x 4 (in-car task blocks) repeated measures ANOVA was performed to analyze the data. There was a clear effect of location, $F(3, 54) = 10.68$, $p < 0.001$. The frequency of correct answers was largest for the middle lane and lowest for the left lane behind the driver. The in-car tasks did not differ significantly. But there was a significant interaction between location and in-car task, $F(9, 162) = 2.26$, $p = 0.02$. This was due to the differences between tasks that were much greater at the locations behind and right in front of the driver than for the locations in the middle and left in front of the driver. The results for these two positions are shown in Table 2. Inspecting the results for the rear left position, one can see that in agreement with the results of Experiment 1 the accuracies for the speech interaction and the PDA writing task were reduced compared to the baseline condition. Accuracy was not reduced for the listening task as was expected from the results in Experiment 1. The results for the front right position in Experiment 1 predicted that the accuracy for the PDA writing task should be reduced because of its visual demand. But this was not the case.

Table 2. Accuracy of answers to the questions about the number of cars in the two lanes for the different in-car tasks

	Right front lane	Rear left lane
Baseline	.75	.72
Listening	.70	.74
Speech interaction	.62	.55
PDA writing	.75	.58

3 Summary and Conclusions

Two experiments were described that tested predictions about the effects of in-car tasks on situation awareness. These predictions were derived from a comprehension based model of situation awareness [4]. In the first experiment an experimental paradigm was tested and used to evaluate the visual and cognitive demands of potential in-car tasks. This paradigm is based on the use of a 1-back task that has to be performed concurrently with an in-car task to measure the visual and cognitive demands of this in-car task. We argued that the detection rate for the 1-back task stimuli depends mainly on the visual demand of the concurrently performed in-car task. The hit rate, that is the proportion of correct responses of all responses made, should be sensitive to the cognitive demand of the in-car task. Consistent with our expectations we found a decreased detection rate for the visually demanding task and a decreased hit rate for the cognitively demanding tasks.

In the second experiment the effect of these tasks on drivers' situation awareness was evaluated in a driving simulator study. Participants had to perform the same tasks that were used in Experiment 1 while driving in a driving simulator. Situation awareness was measured by interrupting the simulation and asking participants about the number of cars on the surrounding lanes.

Contrary to our expectations we found that the PDA writing task that showed a significant decrement in the detection rate in the first experiment led to no decrement in accuracy when the driver was asked for the number of cars in front – irrespective of the lane. It might be that participants were able to use peripheral vision and short glances to watch the traffic situation in front of them while performing the PDA writing task (cf. [7], [8]). In agreement with the model predictions we found that those tasks that were evaluated as highly visually or cognitively demanding, the PDA writing resp. the speech interaction task, lead to more errors in recalling the number of cars at the rear position. The visual demand of the PDA writing task might have reduced the number of glances to the mirrors leading to a reduced integration of information about the rear situation into the situation model. The cognitive load of the speech interaction task presumably led to an increased forgetting of encoded information. Another possibility could be that the cognitive demand of the task led drivers to reduce watching the mirrors and concentrate their glances on a region straight ahead. Therefore, information about the rear situation was gathered to a lesser extent. Such a type of reaction to high cognitive load while driving has been found in other studies on the effect of cognitive demand on driving behavior (e.g, [9]).

The results of the two experiments demonstrate that a task such as the n-back task might be suitable as a measure of the cognitive and the visual demands of tasks. But they also demonstrate that the effects of visual and cognitive task demands on measures of situation awareness while driving are complex and sometimes counterintuitive. These effects have to be examined in more detail using pseudo-in-car tasks with a better controlled profile of visual and cognitive demands.

References

1. Endsley, M.R.: Towards a theory of situation awareness in dynamic systems. Hum Factors 37, 32–64 (1995)
2. Adams, M.J., Tenney, Y.J., Pew, R.W.: Situation awareness and the cognitive management of complex systems. Hum Factors 37, 85–104 (1995)
3. Bailly, B., Bellet, T., Goupil, C.: Driver's mental representations: Experimental study and training perspectives. In: Dorn, L. (ed.): Driver behavior and training. Ashgate, Aldershot (2003)
4. Baumann, M., Krems, J.F.: Situation awareness and driving: A cognitive model. In: Cacciabue, C., Re, C. (eds.): Critical issues in advanced automotive systems and human-centred design. Springer, London (in press)
5. Kintsch, W.: Comprehension: A paradigm for cognition. Cambridge University Press, New York (1998)
6. Baumann, M., Rösler, D., Jahn, G., Krems, J.F.: Assessing driver distraction using occlusion method and peripheral detection task. In: Strasser, H., Kluth, K., Rausch, H., Bubb, H. (eds.) Quality of work and products in enterprises of the future, pp. 53–56. Ergonomia Verlag, Stuttgart (2003)
7. Chiang, D.P., Brooks, A.M., Weir, D.H.: On the highway measures of driver glance behaviour with an example automobile navigation system. Appl. Ergon. 35, 215–224 (2004)
8. Wierwille, W.: Visual and manual demands of in-car controls and displays. In: Peacock, B., Karwowski, W. (eds.): Automotive ergonomics, Taylor and Francis, London (1993)
9. Victor, T.W., Harbluk, J.L., Engström, J.A.: Sensitivity of eye-movement measures to in-vehicle task difficulty. Transport Res. F 8, 167–190 (2005)

Theoretical and Methodological Considerations in the Comparison of Performance and Physiological Measures of Mental Workload

Julien Cegarra[1] and Aline Chevalier[2]

[1] CLLE-LTC Laboratoire Travail et Cognition (UMR CNRS 5263), Centre Universitaire Jean-François Champollion, Place de Verdun, 81012 Albi cedex 9, France
[2] Cognitive Processes and Interactive Behaviours Laboratory, University of Paris X-Nanterre, 200 avenue de la République, 92001 Nanterre cedex, France
Julien.Cegarra@univ-jfc.fr, Aline.Chevalier@u-paris10.fr

Abstract. Mental workload is a central concept to a range of disciplines (including cognitive ergonomics) and has been the subject of various debates. As a result, a number of different techniques have been used by researchers. In this study, we focus on the comparison of performance and physiological measures of mental workload. We describe a puzzle-solving experiment in which we combined multiple measures of mental workload. The results highlight a number of issues. We stress several theoretical and methodological questions that may help to reach a clearer understanding of the nature of the mental workload concept.

Keywords: mental workload; dissociation; dual-task paradigm; eye dilation.

1 Introduction

Mental workload is a central concept to a range of disciplines. For example, in cognitive ergonomics there is a need to regulate the human operator's mental workload in order to prevent both underload and overload, whilst maintaining an adequate level of performance. At the same time, the diverse theoretical foundations of mental workload are still under debate. As a result, a number of different techniques have been used by researchers. Methods for measuring mental workload are usually classified using antitheses, such as direct *versus* indirect, objective *versus* subjective or analytical *versus* empirical. One of the most intuitive classifications focuses on the means for assessing mental workload, taking into consideration performance measures (e.g., dual-task paradigm), subjective measures (e.g., rating scales) and physiological measures (e.g., pupil dilation). Relatively few researchers have attempted to compare methods from different categories (but see, for instance, [7]). Here, we suggest some theoretical and methodological considerations in the comparison of performance and physiological measures.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 264–268, 2007.

2 Performance and Physiological Measures of Mental Workload

2.1 Performance Measures

Many studies indicate a strong link between poor performance in the task and the mental workload experienced by the participant. However, this approach is not always reliable as the relationship between mental workload and performance also depends on task demands, human strategies and individual differences (e.g., [2]). Therefore, different researchers suggested to measure the workload from an additional task. The most frequently used approach to additional task measurement is the dual-task paradigm: Participants are asked to complete a task and at various times are interrupted by a signal (e.g., a tone), which they have to respond as rapidly as possible. It is assumed that the amount of time necessary to respond to the signal reflects the amount of mental workload allocated to the main task. Although some researchers have challenged the assumptions made by this method (e.g., [4]), many others have gone on to use it successfully in a range of tasks, including design activities or information-searching (e.g., [3]).

2.2 Physiological Measures

Physiological measures (such as the measurement of variations in heartbeat, pupil dilation, blink or respiration rate) are based on the assertion that physiological variables reflect changes in cognitive functioning. In this way, the pupil dilation technique is considered to be very sensitive in many tasks and is, therefore, seen to be one of the most relevant physiological measure [1]. Moreover, pupil dilation is considered precise enough to assess subtle variations in workload. However, researchers noted the lack of selectivity of physiological measures (e.g., luminosity is also known to be a factor determining pupil size). In this study, we are comparing this technique to a performance measure, which is more selective but with a lower bandwidth.

3 Method

We designed an experiment based on the Sudoku puzzle (see Fig 1). The aim is to fill a 9 by 9 grid with numerical digits so that every row, every column and every 3 by 3 box contains the digits from 1 to 9. Each participant had to solve 2 puzzles with 2 levels of complexity. During puzzle-solving, we recorded the participant's pupil dilation using the SMI iViewX head-mounted eye-tracking system with a 50 Hz sampling rate. The ambient light was controlled during the experiment, which took place in a noise-attenuated room. As the participant performed the task, we recorded reaction times to standard tones played through the computer's speakers at 10 to 20 second intervals. The intrusiveness of the signal range was controlled in a pilot study.

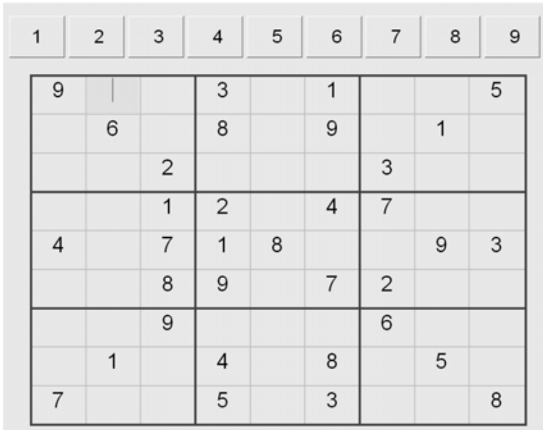


Fig. 1. The Sudoku puzzle: The participant has to fill the grid, selecting the numerical digits so that every row, every column and every 3 by 3 box contains the digits from 1 to 9

4 Results

To precisely compare the two techniques, we focus the analysis on the easiest of the two puzzles on one participant. Figure 2 displays two curves along a temporal axis: (1) the reaction times on a scale from 0 to 500 msecs, displayed as a curve (rather than discrete plots) in order to facilitate comparison; (2) pupil diameter on a scale from 3 to 4.5 mm, as well as a smoothed curve of these values.

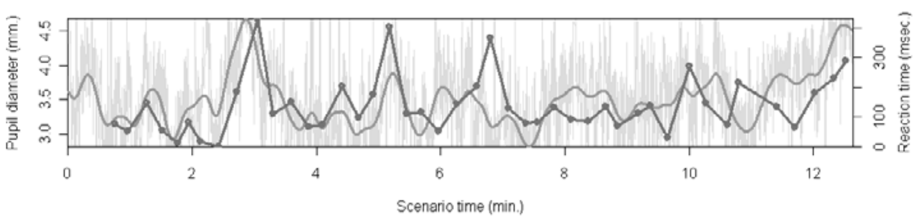


Fig. 2. Curves of the reaction times (msec.) and pupil dilation measures (mm.) for the easiest puzzle

A comparison between the two curves cannot be made on exact values as the unit measurements differ for each scale. However, the shape of the curves provides relevant information. On closer examination of the two curves, differences in shape become apparent:

- The peaks do not appear exactly at the same moment. The peaks in the pupil dilation curve sometimes precede (e.g., at about 3 minutes on the x-axis), seem to just follow (e.g., at about 5 minutes) or seem unrelated (e.g., at about 7 minutes) to a corresponding peak in the reaction time curve.

– The physiological measure captures more changes. Even while looking at the smoothed curve of pupil dilation, the curve of the reaction time often badly approximates its shape. So the sensitivity of the curve is indeed dependent on the sampling of the measure.

5 Discussion

These results illustrate the benefits of comparing physiological and performance measures of mental workload.

From a methodological point of view, it is almost impossible to discuss the changes in mental workload for the reaction times without using another measure. For example, care must be observed when claiming that an increase in workload is followed by a decrease (or vice versa) because the shape of the curve is uncertain between the known values. Another consideration results from the diagnosticity of the eye dilation measurement technique. As Wickens and Hollands [8] stated, physiological measures have to be considered within the context of separate pools of resources. The apparent differences in the peaks between the eye dilation measurement technique and the dual-task paradigm may result from the fact than the eye dilation may be more diagnostic of the load being considered. As noted by Wilson and O'Donnell [9], an increased load does not always result in increased overall activation. The more diagnostic measure would detect changes in the load, whereas the general measure of load would not. This would also mean that even if the changes in the two graphs seem to imply a close correlation, one has to be careful interpreting the results as the changes do not necessarily have a common cause.

Moreover, from a theoretical point of view, questions arise concerning the validity of the reaction times to measure workload. It is necessary to distinguish between delays in reaction times which are related to workload and delays related to conflicting attentional demands or psychological refractory period (i.e. the signal is appearing when a response to the main task is underway). A more detailed comparison of physiological measures and the dual-task paradigm (with a larger participant sample), as suggested by Isreal et al. [6], might provide more details about the theoretical foundations of the dual-task paradigm. This may take the form of a comparison between the dual-task paradigm to the irrelevant-probe technique. This technique involves recording of physiological measures (such as ERP) in which, unlike the dual-task paradigm, the participant has to ignore the probes. Several authors have noted that attention not focused elsewhere is attracted by the probes even if the participant does not directly pay attention to the probe [5].

6 Conclusion

These results stress that such analysis of performance and physiological measures are crucial, especially since many studies measure mental workload with an additional task. For example, mental workload in car driving is evaluated whilst receiving and holding conversations on a mobile phone (i.e. measuring gaps in the telephone conversation).

Therefore, more studies are necessary to precisely identify the theoretical and methodological limitations of performance and physiological measures of mental workload.

References

1. Beatty, J.: Task evoked pupillary responses, processing load, and the structure of processing resources. Psych. Bull. 91, 276–292 (1982)
2. Cegarra, J., Hoc, J.M.: Cognitive Styles as an Explanation of Experts' Individual Differences: A Case Study in Computer-Assisted Troubleshooting Diagnosis. Int. J. of Hum.-Comp. Stud. 64, 123–136 (2006)
3. Chevalier, A., Kicka, M.: Web designers and web users: Influence of the ergonomic quality of the web site on the information search task. Int. J. of Hum.-Comp. Stud. 64, 1031–1048 (2006)
4. Fisk, A.D., Derrick, W.L., Schneider, W.: A Methodological Assessment and Evaluation of Dual-Task Paradigms. Cur. Psych. Res. & Rev. 5, 315–327 (1987)
5. Kramer, A.F., Trejo, L.J., Humphrey, D.: Assessment of mental workload with task-irrelevant auditory probes. Biol. Psych. 40, 83–100 (1995)
6. Isreal, J.B., Chesney, G.L., Wickens, C.D., Donchin, E.: P300 and Tracking Difficulty: Evidence for Multiple Resources in Dual-Task Performance. Psychophysiology 17, 259–273 (1980)
7. Miyake, S.: Multivariate workload evaluation combining physiological and subjective measures. Int. J. of Psychophysiology 40, 233–238 (2001)
8. Wickens, C.D., Hollands, J.: Engineering psychology and human performance. Longmans, London, UK (2000)
9. Wilson, G.F., O'Donnell, R.D.: Measurement of operator workload with the neuropsychological workload test battery. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 63–100. Elsevier, Amsterdam (1988)

Results of a Tailored Communication Framework Through E-Health

Eva del Hoyo-Barbolla, Emanuele Carisio, Marta Ortega-Portillo, and María Teresa Arredondo

Life Supporting Technologies, ETSI Telecomunicación, Ciudad Universitaria s/n, 28040 Madrid, Spain
{evahb,ecarisio,mortega,mta}@lst.tfo.upm.es

Abstract. This paper proposes a new framework to enhance the interaction of individuals with e-health and encourage the adoption of healthy lifestyles. Achieving a lifestyle change by the use of e-health is a complex issue that can be broadly addressed by analysing, in parallel the individuals' attitude towards their health condition and their approach and readiness to monitor and change their attitude by the use of new technologies. Our work has been to conceive, develop and evaluate a novel framework that explains how to promote the acceptance of e-health in terms of the perception of healthcare and of the use of technology to perform a desired or recommended lifestyle change. In order to place the user at each dimension a set of questionnaires was designed and implemented. These questionnaires assisted us in understanding what personalised information needs to be provided according to the stage the user is at as well as to other variables (such as age, cultural background, etc). Moreover, we created a communication strategy to provide tailored information regarding promotion and prevention of healthcare by means of e-health and promote the use of technology solutions to improve individuals' habits by enhancing their interaction with technological means. The framework was finally evaluated by users and healthcare professionals.

Keywords: E-health acceptance, healthcare promotion, e-health communication strategies.

1 Introduction

Preventing or delaying illness and death from chronic disease is possible [1]. Many of these diseases can be prevented or ameliorated through behaviour changes [2]. At least 80% of all cardiovascular diseases and "type 2" diabetes and over 40% of cancer could be avoided through healthy diet, regular physical activity and avoidance of tobacco use, which are behaviours that can be influenced and modified through education.

ICT and its applications are increasingly looked upon as a potential answer to the requirements of a modern society, with demands for better healthcare, improvements in medical outcomes, and maintenance of a relatively high quality of life, especially

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 269–278, 2007.

with the onset of chronic health conditions coming to the fore as a key issue. Furthermore, implying a view of utilising the technology as a tool for (re) addressing the prevailing state of affairs, ICT tools and applications are also seen as having a potential to support an enhanced access to health information in general and indeed, to the health system itself in particular [3]. Characteristics such as the tailoring of messages, instantaneous feedback, appeal or engagement are potential advantages that new ICT can provide (adapted from [4]) and that may be of enormous benefit to attain behaviour change.

ICT innovations have the capacity to support empowerment of individuals in managing their health concerns and acquiring the necessary resources to achieve their goals. Moreover, it can ensure ubiquitous availability of the tools and communication channels necessary to support empowerment [5]. However, although there is much emphasis on the adoption of ICT implementation in the field of healthcare (e-health) [6], [7] there is still a lack of understanding the in-depth rationale to how these applications assist individuals to effectively change lifestyle behaviours.

On the one hand, there is a scarce on studies that relate the use of ICT to change attitude towards healthcare and on the other, the studies often focus on the technical aspects of ICT failure, neglecting what has been learned from the behavioural sciences about humans and their interaction with ICT; this may be the answer to understand the added value that ICT bring over traditional channels of healthcare provision. The success of ICT implementation in order to cause a change in individual's behaviour is often grounded in behavioural science, using theories and models to identify conditions and determinants of successful use [8]. Therefore, working out why an individual would use an e-health application to change his/her lifestyle means paying attention both at the behavioural aspects both in attitude towards healthcare and in technology readiness. This paper proposes a new framework that provides a tailored communication strategy by means of e-health promotion applications to understand the reasons of individuals to use new ICT to perform changes in their lifestyle. This is a multifold issue that can be tackled broadly by analysing their willingness to change health behaviours and to do so by using an e-health application.

Besides, this paper presents a practical implementation of the framework used for encouraging healthy eating, implemented using "portlets", a novel Web component based on Java technology that returns dynamic content to user requests [9], allowing the highest possible degree of personalization. The use of new web design technologies increases also the system's dynamism and interactivity.

2 Methods

Our intention is to build a framework that provides personalised information according to a set of variables relevant to the user. Tailoring both the channel and the content of the message has proven essential for persuading individuals to change their health behaviour [10]. In order to personalise information, a 5-step approach was defined based on the modification of a strategy for the creation of a personalised healthcare communication strategy [11]:

- Step 1: Analysing the problem to be addressed and understanding its determinants and defining a new model

- Step 2: Developing an assessment tool to measure a person's status on these determinants
- Step 3: Creating tailored messages that address individual validation of determinants of the problem
- Step 4: Developing algorithms to link responses from the assessment into specific tailored messages to create the final health communication.
- Step 5: Final evaluation.

A description of these steps and how they have been applied to the creation and implementation of the framework follows:

- *Step 1*

Theoretical models are useful in predicting which patients will use e-health and in understanding what factors influence their decisions. Models also can aid in designing and evaluating the ability of specific e-health applications [12]. Currently there is no integrated framework that includes a sufficiently broad set of influencing factors to understand the multidimensionality of the reasons why people use ICT to embrace a healthcare change [8]. There have been previous attempts in understanding whether a single factor may have an effect on their e-health adoption, by partially applying some of the psychological models that explain behaviour change. However, it was noticed that applying only one of the theories (i.e. the "Stages of Change") [13] was somehow limited as there are a wider variety of factors that influence a user's decision to adopt a change in his/her life in order to perform a healthcare change.

Moreover, without addressing the full range of factors, strategies to change usage behaviour run the risk of being ineffective because they fail to recognise interdependencies between individual and organizational factors [8]. Our research focuses on how to promote effective change in health behaviour by means of modern ICT. Therefore, a detailed study focusing on both healthcare and ICT dimensions was carried out to understand the stage the user is at.

In the first case research developed in psychology and public health that attempts to understand the promotion of health among the populations is complex and there are a number of significant theories and models that underpin the practice of health promotion and individual's attitudes towards the change, these being mainly: the Health Belief Model [14], Theory of Reasoned Action [15] and the Theory of Planned Behaviour [16], the Trans-theoretical (stages of change) Model [17] or the Precaution Adoption process Model [18]. These theories explain health behaviour change by focusing on the individual with the principal intention of providing information either to improve knowledge or change behaviour.

On the other hand, our task is to understand why users would make use of an ICT platform (denominated from now on "e-health" platform) in order to perform a change in their attitude towards health (i.e. use the Internet or a mobile phone to help them quit smoking, to encourage them to follow a diet, etc). However, we only found adoption theories that explained why a user would make use of an ICT platform to carry out a specific ICT-work related task (i.e. adoption of spreadsheets, use of email, word processor, etc) in different environments (companies, hospitals, government agencies, etc) and situations, although not health related attitudes. Studies elaborated show the interest of individuals in using an ICT application but there is no holistic framework to explain the link between these last two (i.e. the use of ICT to perform

changes in health behaviour). Our work has been to conceive and develop an integrative framework that explains the different stages the user is at both in terms of the perception of healthcare and the use of technology to perform any change.

- *Step 2*

In order to understand what the individual's status is regarding the variables previously explained, we designed and implemented a set of questionnaires to place the user at a stage in each of the paths so that we understand what personalised information needs to be provided according to the stage the user is at. Other factors (personality, age, etc) need to be taken into account but in this work we will focus on the stage at the healthcare and the ICT level.

- *Step 3*

Working together with the health professionals we modelled the structure of a medical intervention in the field for an individual placed at each of the stages in the different pathologies studied (6 month healthy-eating plan and 10000-step programme for diabetic patients), arriving to the construct of personalised messages for each case.

- *Step 4*

After having modelled the intervention we modelled the tool for the healthy-eating plan that provides the personalised messages according to the user needs. This tool is a Web-based tool in which messages are provided through different channels according to different variables. This solution has an additional advantage, since it makes such information available to the user through the Internet. For this purpose, different innovative technologies relative to dynamic Web development have been analysed, such as Java Portlet [9], JSP, Oracle databases, HTML or Macromedia Flash. The selection of these technologies is based on the following requirements: interactivity, personalisation, portability and usability, as well as a potential seamless integration in any kind of system. The whole process described in step 4 consists itself of five different stages. In each stage, different technologies have been selected to achieve the required functionalities of the full system. These stages are: Login, Questionnaires, Personalisation, Storage and Presentation.

At the Login stage the user's individual information is recalled from the database records, which have been previously stored in the general profiling process. The Questionnaires about health behaviour has been developed. Their functionality is to define the users' main concern and to discover their motivation status regarding their health [19] and their attitude towards new ICT. The previous steps have been already used to provide required methodologies to classify the user into each stage, and to suggest appropriate motivational techniques for them.

The Personalisation stage adapts the information to be delivered to the user. The user's profile is modified after the questionnaires are filled in. Thus, the system is personalised according to the user's preferences and completed with the most suitable information.

The Storage stage responds to the need of storing the users' profiles, with all the corresponding security and coherence requirements, to be later used all along the session. Finally, at the Presentation stage the personalised information and recommendations according to the results of the previous stages are presented to the user. The information shown has been selected in collaboration with health professionals.

To model the information provided, the structure of a standard medical intervention for a user suffering any of the different pathologies studied has been modelled with the guidance of healthcare professional.

The final communication strategy was designed and tested with the doctors.

- *Step 5*

An evaluation of the framework is already being carried out. Results show promising and the feedback obtained from the different stakeholders is currently being used to perform a second iteration of the whole communication methodology.

3 Results

3.1 The Model Proposed

A model that takes into account in parallel both the healthcare stage of the individual as well as his/her technical readiness to perform a change using technologies was conceived and designed. The added value of this model is that it allows placing the user at a stage in both dimensions (see Figure 1).

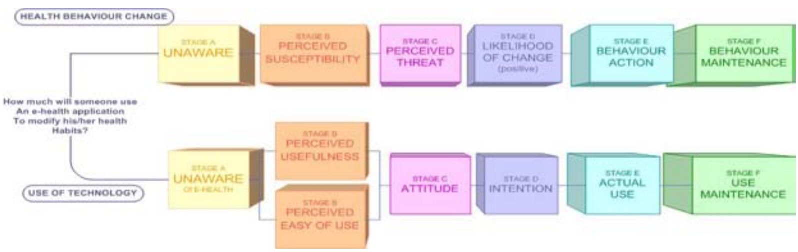


Fig. 1. The model proposed

As there are many similarities across the different theories, the upper path of the model (health behaviour change) is an adaptation from an integrative model proposed by Fishbein [20]. The model proposed considers the key variables of the models previously cited [4]. This path is appealing as it reflects the importance of intra-individual, environmental factors and self-efficacy. The main stages of the model are:

- Unawareness: an individual never heard about the issue and therefore may have no opinion about it.
- Perceived susceptibility: One's opinion of chances of getting a condition
- Perceived threat: the belief that one is susceptible to a specific problem
- Self efficacy: the belief that one has the ability to change one's behaviour
- Attitude to change: making a plan to change behaviour
- Action: implementing the plan to change behaviour
- Maintenance: continuation of behaviour change

This model supposes that individual’s usually pass through the different stages, although the pace may be different for different individuals or different behaviours.

Movement backward toward an earlier stage is also possible (although not to stage 1 obviously). The lower path of the model is an adaptation of the Technology Acceptance Model (TAM) model proposed by Davis [21], [22] that models how users come to accept and use technology. Its stages are similar to the health behaviour model, although adapted to technology. The difference of the stages lies in these two stages [22]:

- Perceived usefulness: degree to which someone believes that a system would enhance the job performance.
- Perceived ease-of-use: degree to which a person believes that using a system would be free from effort.

3.2 The Questionnaires Designed

In order to be able to place the user in a stage a questionnaire/survey instrument was developed based on the model presented. It took into account the substantive factors influencing the adoption and readiness of individuals to perform a health behaviour change and in parallel to do it by means of an e-health application. Based on the stages identified in the model, two different sets of questions were developed, one for each branch. The questions try to find out the attitude of the user towards the two specific variables: one related to the interest in health (how the user relates to the healthcare stage) and the other related to their motivation and skills to use ICT to promote their well-being (whether the user would be ready to use an e-health application).

Although both stages are linked, two different questionnaires are presented to the user. Questionnaires are implemented in JSP, to allow conducting the user "intelligently" to the most appropriate question depending on previous answers. There are several ways to arrive to the same stage but with different characteristics depending on the answers to key questions and these are also taken into account when tailoring the information. The e-health questionnaire identifies whether users have knowledge, motivation or access, to use ICT to make a health change.

The health questionnaire (see Figure 2) assesses whether users have the awareness and the right guidelines to control their health, it they feel susceptible to suffer from that pathology and if users are following any healthcare plan. With these objectives in mind, three specific strategies to encourage self-care can be followed: offering correct and personalised information, helping users use this information and promoting ICT as a means to support healthcare by easing access to quality information and specific services, helping health problems follow-up and making the relation between patients/users and health professionals closer.

3.3 The Intervention Selected According to Stage

Once users complete both questionnaires they are allocated to a stage in their attitude towards the healthcare and the use of e-health to achieve a positive change. The process that models the intervention followed by the professional is selected.

3.4 The Algorithm Developed

The Web platform described has been implemented with the following features:

- To offer personalised information in form, tone and content.
- To be visual, dynamic and interactive.
- To be able to be integrated in an e-learning system.

In this sense, a training activity consists of a multimedia session through the Internet and presented by means of a visual and friendly interface. In particular, the implemented tool is a portlet using many of the characteristics of the specification JSR 168 [9], providing the user a two modes portlet: VIEW and HELP. The VIEW mode allows the user to get authenticated. After that, users are asked about their main concern and the questionnaire is displayed to classify users.

Once the user has been classified the information is stored in the database and personalised information is presented in terms of form, tone and content. The tailoring process is performed by using techniques that, using the adopted methodology [19], are the most appropriate to instruct, to convince and to motivate users. Users' awareness to take control and be co-responsible of their healthcare, is raised by presenting only relevant information through a personalised interface.

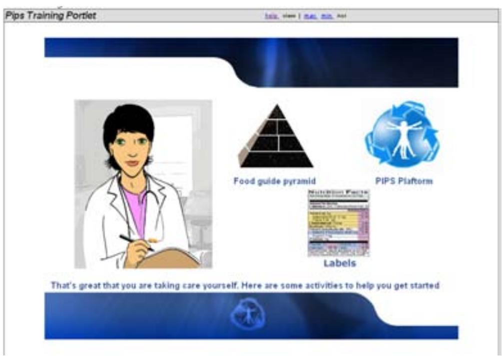


Fig. 2. Personalized information presented to the user

The database structure offers different types of information: profile, lifestyle characteristics and behaviour towards health and e-health status. The latter is the most important since it can be dynamically changed every time a user accesses the system and modifies any aspect related to his health behaviour. In this way, if the user has not changed the information introduced in previous sessions, the results displayed by the system reflect the user's needs and motivation. Information in the database must observe confidentiality and security requirements. All these requirements of structure, integrity and coherence suggested the use of a well known and established relational database such as Oracle.

3.5 Personalised Information

According to these stages, the information offered to the user is selected in line with the main features identified, and presented in the most suitable way so that a change

to improve his/her health with the use of ICT can be achieved. Figure 6 shows the information for a user in Action stage (this is related to obesity as a pathology).

3.6 Evaluation of the Framework

The survey developed for the evaluation of the framework concept among the general population and non-specialists healthcare professionals was responded by 20 people. Healthcare professionals belief personalised e-health can help patients and the overall population towards improving the effectiveness of people and perform tasks in a more effective way. The system is believed to be useful in easing people's routine (i.e. reminders, suggestions, feedback, etc), and to allow do tasks more effectively, such as taking a routine to take care of one's healthcare. On the other hand the less valued aspect is the perceived capabilities of the system to persuade and encourage self-care. In this regard and in order to improve this factor, this research worked intensively on the enhancement of this functionality by introducing motivating factors that encourage people to use e-health systems to improve the care of their healthcare.

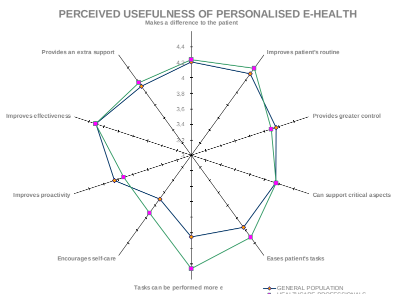


Fig. 3. Results of expert evaluation of the concept of the framework

4 Discussion and Conclusions

Information is critical to health-related decisions and the way this information is selected and processed plays a pivotal role in the decision making of individuals. This model allows the classification of users according to their attitude towards healthcare, and their readiness to adopt the change by the use of new technologies. It provides a framework for the provision of personalised information according to these and other key variables. Besides, the application of this framework promotes the empowerment of the individuals, as well as guidance, monitoring, through ICT and it will certainly make an impact on health-related behaviour.

This framework is novel as it proposes a framework that may be applied to the conception, design and evaluation of any e-health application. Besides, it aims at providing a communication structure for different application targets (medical informatics, public health informatics, etc) and to different stakeholders (healthy individuals, patients, professionals, etc). The framework promotes empowerment of the individuals by providing tailored information, as well as the acceptance of technology and it will certainly make an impact on health-related behaviour. Overall, this framework is likely to

provide deeper insights into the process of improving e-health acceptance so it can meet ongoing individuals' needs and become an increasingly integrated part of the preventive healthcare services value chain.

This framework allows enhancing the user modelling process by taking into account both health behaviour aspects as well as technological, not having been considered up to this moment and could be part of the explanation of e-health common underuse. It may assist both healthcare professionals and individuals to have a deeper understanding about both, the provision of healthcare and the delivery channel. However, the inclusion of other dimensions is also necessary in order to be able to effectively reach the individuals.

These tools are not meant to substitute the role of the professional, but help patients in raising awareness of their conditions and encouraging them to assume healthier lifestyles. This tool serves as a guide through the different stages of motivation described in the e-health framework until the desired healthy behaviour is adopted. Personalisation and usability are characteristics that allow this tool to create a user-centred healthcare model. Their needs and preferences are considered to allocate healthy habits promotion.

Other advantage of this tool is its portability, due to the use of Portlets, implemented as Web files (.war) that can be deployed by a portlet container being, at the same time, their content independent from it; the tool developed can be used in any other platform without the need of introducing any modification.

Overall, this framework is likely to provide deeper insights into the process of improving the acceptance of e-health, so it can meet ongoing individuals' needs and become an increasingly valued part of health care services.

Acknowledgments. We would like to thank the PIPS project (partially funded by the EC, IST 507019) in which this framework will be tested.

References

1. World Health Organisation. Facing the Facts: solving the chronic disease problem (2005)
2. U. S Department of Health and Human Services. The Health Benefits of Smoking Cessation: centers for Disease Control, Centers for chronic Disease Prevention and Health Promotion, Office on Smoking and Health; Report N. (CDC) 90-8416 (1990)
3. The Euser Project. eUser – Workpackage 1: Conceptual and Analytical Framework. D1.1: eUSER Conceptual and Analytical Framework (Part A and Part C). European Commission, 6th Framework Programme. Contract number: IST- 2002-507180
4. Glanz, K., Rimer, B., Lewis, F.M.: Health Behaviour and Health Education: Theory, Research and Practice, 3rd edn. John Wiley & Sons, Chichester (2002) ISBN 0-7879-5715-1
5. Brennan, P.F., Safran, C.: Empowered Consumers. In: Lewis, D., Eyenbach, G., Kukafka, R., et al. (eds.) Consumer Health informatics, Springer, Heidelberg (2005) ISBN 0-387-23991-x.
6. Lorenzi, N.M., Riley, R.T., Blyth, A.J., Southon, G., Dixon, B.G.: Antecedents of the people and organizational aspects of medical informatics: review of the literature. J. Am. Med. Inform. Assoc. 4, 79–93 (1977)
7. Curtis, B., Krasner, H., Iscoe, N.: A field study of the software design process for large systems. Commun. ACM 31, 1268–1286 (1988)

8. Kukafka, R., Johnson, S.B., Linfante, A., Allegrante, J.P.: Grounding a new information technology implementation framework in behavioral science: a systematic analysis of the literature on IT use. Journal of Biomedical Informatics 36(3), 218–227 (2003)
9. DeWolf, D.: Introduction to the Java Portlet Specification, (September 2005), URL: http://www.developer.com/java/web/article.php/3547186
10. Kukafka, R.: Tailored Health Communication. In: Lewis, D., Eyenbach, G., Kukafka, R., et al. (eds.) In consumer Health informatics, Springer, Heidelberg (2005) ISBN 0-387-23991-x.
11. Kreuter, M., Farrell, D., Olevich, L., et al.: Tailoring Health Messages: Customising Communication with Computer Technology. Lawrence Erlbaum, Mahwah, NJ (2000)
12. Vance Wilson, E., Lankton, N.K.: Modeling Patients' Acceptance of Provider-delivered E-health. Journal of the American Medical Informatics Association, 11(4) (July / August 2004)
13. Del Hoyo-Barbolla, E., Fernandez, N., Ramirez, C.M., Tortajada, D.: de la Cruz, J.N., Arredondo, M.T.: Personalised training tool using Virtual Reality. In: Proceedings of the 1st workshop on personalisation for e-health. 10th International Conference on User Modelling. Edimburgh, UK (2005)
14. Fishbein, M., Ajzen, I.: Beliefs, Attitudes, Intention, and Behavior: an Introduction to Theory and Research. Addison-Wesley, Reading, MA (1975)
15. Ajzen, I., Fishbein, M.: Understanding Attitudes and Predicting Social Behavior. Prentice-Hall, Englewood Cliffs, NJ (1980)
16. Fishbein, M.: A theory of reasoned action: some applications and implications. In: Page, M.M. (ed.) 1979 Nebraska Symposium on Motivation, University of Nebraska Press, Lincoln, NE (1980)
17. Prochaska, J.O., DiClemente, C.C.: Stages and processes of self-change of smoking: Toward an integrative model of change. J Consult Clin. Psychol 51, 390–595 (1983)
18. Weinstein, N.D., Sandman, P.M.: A model of the Precaution Adoption Process: Evidence from home Radon testing. Health psychology 11, 170–180 (1992)
19. Del Hoyo-Barbolla, E., Arredondo, M.T., Ortega-Portillo, M., Fernández, N., Villalba, E.: A new approach to model the adoption of e-health. In: 13th IEEE Mediterranean Electrotechnical Conference (MELECOM, 16-19 May, Spain. Electronic Proceedings, IEEE Catalog Number 06CH37756C (2006) ISBN 1-4244-0088-0
20. Fishbein, M.: The role of Theory in HIV prevention. The role of Theory in HIV prevention. AIDS care 12(3), 273–278 (2000)
21. Davis, F., Bagozzi, R.P., Warshaw, P.R.: User acceptance of computer technology: A comparison of two theoretical models. Management Science 35, 982–1003 (1989)
22. Davis, F.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 13(3), 319–340 (1989)

Effects of Cognitive Training on Individual Differences in Attention

Jing Feng and Ian Spence

Engineering Psychology Lab, Department of Psychology, University of Toronto,
100 St. George. Street, Toronto
Canada M5S 3G3

Abstract. Selective attention is responsible for detecting, localizing and identifying a target while neglecting distractors [1],[2]. A superior capacity in selective attention contributes to good performance in tasks that require monitoring the environment and searching for a target [2],[3],[4]. Since it is our goal to optimize work efficiency, understanding individual differences in attentional capacity and whether they are mutable is important. Our first experiment demonstrates the existence of systematic individual differences in selective attention. More remarkably, our second experiment shows that appropriate cognitive training using an action video game can alter selective attentional capacity. Furthermore, individuals with the poorest initial scores gain most from the training. We show that these gains cannot be attributed to regression effects alone. Thus we conclude that individual differences in attentional capacity can be reduced or even eliminated by training.

Keywords: engineering psychology, cognitive training, individual differences, selective attention, video games.

1 Introduction

Selective attentional capacity is one of the principal bottlenecks in human information processing [5]. This capacity has a strong impact on performance in real life situations (e.g. driving, computer programming, piloting, power plant control) [3],[4],[6]. Lack of sufficient mental resources for selective attention often results in failure to detect sudden changes in the environment, diminished situational awareness, as well as the reduced ability to identify a target. Since individual differences are known to exist in many areas of cognition, it is important to determine whether there are systematic individual differences in attentional capacity, and more significantly, whether performance can be improved by cognitive training.

1.1 Selective Attention

Due to our limited mental resources, attention is selectively allocated to certain channels of information [4]. One direct consequence of this selection mechanism on our vision system is our limited spatial distribution of attention. This is the visual area

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 279–287, 2007.

from which we are able to extract information within one eye fixation [2],[3],[4]. Another consequence of selection is that, even within the useful field of view, we can apply mental resources only to some visual components (targets) [4]. This activity requires identifying the target and assigning resources for further processing, while rejecting the distractors and prohibiting them from gaining access to the resources [7].

In many occupational situations, selective attention is critical. For example, in an aircraft cockpit or a control console, a great deal of information is often presented at the same time. Selection of information becomes a challenge for the operator. Superior capacity in allocating attentional resources could benefit the operator in many tasks that require extracting and processing important information, as well as detecting early signals of potential problems. In addition, a larger spatial distribution of selective attention could confer another advantage during the performance of those tasks since fewer eye movements would be needed to acquire the same amount of information. For example, several studies have already identified a strong association between the size of the spatial distribution of attention and driving performance [2],[3]. The ability to concentrate and focus on a number of objects over a large visual area, within one fixation, can be helpful in several applied areas.

1.2 Individual Differences

In any profession, certain kinds of cognitive abilities and skills are necessary. As has been suggested by the theory of competence-related beliefs, individuals who do not have adequate required skills are less likely to feel competent, less likely to be successful, and also less likely to hold positive attitudes toward the profession [8],[9],[10]. Thus, individual differences in certain types of cognitive ability contribute to variation in the career choices of individuals and also to subsequent performance in the occupations that require these abilities. For example, the consequences of individual differences in selective attention can be quite considerable in occupational tasks that require constant monitoring of a continuously changing complex visual field (e.g. piloting, driving, and large system monitoring).

There is a large body of literature on individual differences in cognitive abilities. For instance, it is known that individuals differ greatly in their ability to mentally represent and manipulate spatial information [11],[12],[13]. Similarly, there are large individual differences in performing tasks that require memory [14],[15],[16]. Since selective attention is a lower level capacity that is thought to support many of these higher level cognitive functions [17],[18],[19], it is reasonable to expect that significant individual variations in selective attention may contribute to individual differences in higher cognition. Indeed, some suggestive associations have already been noted: for example, young children [20] and older adults [2], whose higher level cognitive skills are generally inferior to those of young adults, also perform more poorly on some attentional tasks. In addition, pathological populations, including Alzheimer’s patients [21] and children with attention-deficit disorder [22], show inferiorities on selective attention tasks compared to healthy populations. However, little attention has been devoted to possible individual differences within the group of normal healthy young adults. Given the fact that young adults constitute a large fraction of the labor force, understanding the possible individual variations in selective attentional capacity becomes important. Moreover, although individual differences in

cognition can been attributed to certain biological causes [23],[24], environmental factors may also play a part. It has been demonstrated that practice on some cognitive tasks can produce substantial improvements in performance [1] and the improvement is greatest in those who have not previously practiced [25]. Hence, environmental influences seem to be capable of making modifications. We are interested in identifying the possible individual differences in selective attention, as well as the degree of modification we are able to make on those differences.

1.3 Cognitive Training

Cognitive training is a widely used technique for improving an individual's performance on certain cognitive tasks using either the cognitive task itself [26] or relevant associated activities, such as virtual reality settings [27] or video games [1],[28]. Video games have been found to have positive effects on some abilities like attention and spatial cognition [1],[7],[28],[29],[30],[31]. The observed improvements have usually been attributed to the exercise of the particular cognitive ability during game playing [28],[31]. The magnitude of prior individual differences in certain cognitive capacities was diminished or even eliminated in some cases [29],[30]. The category of video game that seems to impose the heaviest demands on selective attention is the action video game. This genre of video game has been associated with superior performance on selective attentional tasks [1],[7],[31]. Interestingly, the superior measured performance of video game players can be matched by non-video game players by training with an action video game for as few as ten hours [1]. In the experiments described below, we investigate whether an action video game can contribute to improved individual performance, and whether playing an action game can modify individual differences in selective attention.

2 Experiments

The first experiment explored individual differences in the spatial distribution of selective attention. The second experiment investigated the possibility of improving selective attention using an action video game as a training device. We were particularly interested in determining whether such training could modify any pre-existing individual differences in selective attention.

The method of measuring selective attention was the Useful Field of View (UFOV) task, which assesses the spatial distribution of attention and the ability to pick out a target among distractors. The stimulus display consisted of twenty-four squares with each uniquely localized at an eccentricity of 10^o or 20^o or 30^o in eight equally spaced directions. One location was randomly selected for the target—a filled square surrounded by a circle. The other 23 locations were left for the distractors, which were unfilled squares. On each trial, the participants were required to report the direction of the target. The whole experiment consisted of 240 trials. A sample trial is shown in Fig. 1.

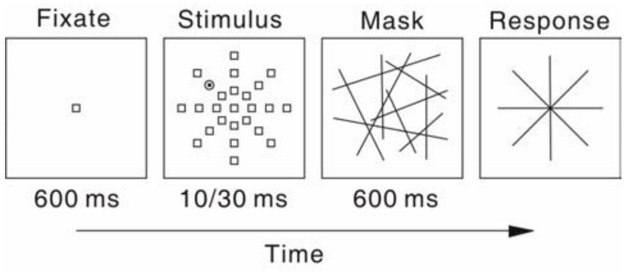


Fig. 1. The Useful Field of View (UFOV) Task

2.1 Experiment 1

Forty-eight undergraduates participated and were balanced with respect to gender (half males and half females), action video gaming experience (half video game players vs. half non-video game players—VGP vs. NVGP), and major (half arts and humanities majors vs. half science and engineering majors), thus forming eight groups of six participants. Each participant completed the UFOV task.

A 2 (gaming experience) x 2 (gender) x 2 (major) x 3 (eccentricity) ANOVA was used to investigate the group differences. In performance, VGPs made significantly fewer errors than NVGPs (gaming experience: $F(1,40) = 34.381$, $p < .001$). A gender difference was found with males outperforming females (gender: $F(1,40) = 5.027$, $p = .031$). The disparity was largest at small eccentricities (gender: eccentricity=10^{o}, $F(1,40) = 6.190$, $p = .017$). Interestingly, the overall difference between males and females is largely attributable to the significant gender difference among NVGPs (significant gender difference among NVGPs: Mean Difference (Std. Error) =.121(.048), $p = .015$; non-significant gender difference among VGPs: Mean Difference (Std. Error) = .028(.048), $p = .564$). In other words, while both males and females in the VGP group performed well, the NVGPs had much higher error rates, with the females performing much more poorly than the males (Fig. 2a). In addition, on average, the science and engineering majors outperformed the arts and humanities majors (major: $F(1,40) = 6.991$, $p = .012$) which implies that students with superior selective attentional capacity are more likely to choose programs in science and engineering (Fig. 2b).

The results of Experiment 1 are suggestive. Since no gender difference was found in the VGP group, playing action video games may benefit females and males equally (Figure 2a). In other words, playing action video games may have the effect of reducing or eliminating prior gender differences in selective attention. On the other hand, these results might simply reflect a selection bias: Individuals, of either sex, with superior attentional skills may have chosen to play action video games whereas others with lesser capacities may have avoided them. However, if attentional capacity can be improved by training NVGPs using action video games, this would weaken the criticism that the observed differences in performance in Experiment 1 are simply the result of selection biases. Our Experiment 2 was designed to test the hypothesis that playing an action video game can improve an individual's attentional capacity.

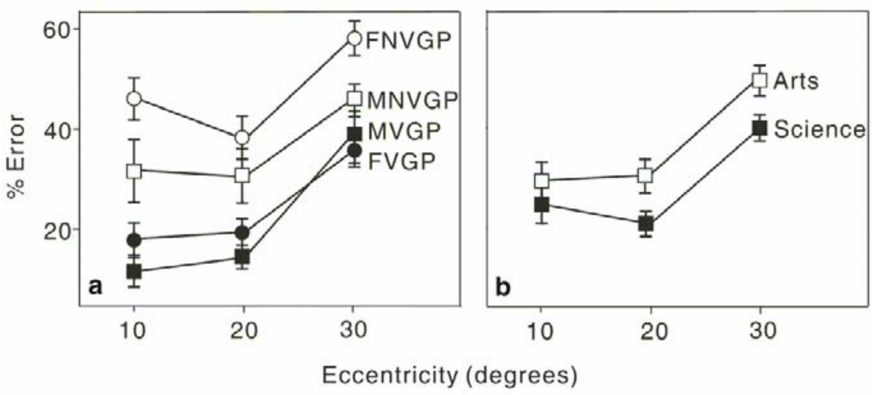


Fig. 2. Performance (percentage errors) of different groups at three eccentricities: a) male video game players (MVGP), female video game players (FVGP), male non-video game players (MNVGP) and female non-video game players (FNVGP); b) performance of Arts and Humanities majors (Arts) and Science and Engineering majors (Science)

2.2 Experiment 2

Twenty undergraduates (six males and fourteen females) without action-video-gaming experience participated. Each participant received a pre-test on the UFOV task, ten hours of training on a video game and then a post-test on the UFOV task. In order to separate the effects of selective attention from the visuo-motor skills required to play video games in general, we used a control group that was trained with a video game which did not require any special ability in selective attention. Males and females were divided equally into two groups with matched pre-test performances on the UFOV task. During the training phase, participants in the experimental group were required to play a classic action video game, *Medal of Honour: Pacific Assault.* Meanwhile, the control group was trained with another 3D game, *Ballance,* which requires fine hand-eye coordination but not superior selective attentional capacities. The ten hours of training for each participant was conducted in several sessions of one to two hours within a maximum of three weeks in the lab.

A 2 (gender) x 2 (training) ANOVA was used to explore group differences. Before training, the performances of the experimental group and the control group on the UFOV task were very similar. Of course, this was to be expected, since we assigned participants to the two groups so that the pre-training group averages would be as close as possible. However, on the post-test, significant improvements on the UFOV task were observed in the experimental group but not in the control group (Fig. 3). Participants who had trained with the action video game made significantly fewer errors at all eccentricities (training: $F(1,8) = 284.171$, $p < .001$). There was no difference between pre-test and post-test performance in the group that had been trained with the non-action video game (training: $F(1,8) = 1.143$, $p = .316$). This result suggests that, compared to a non-action video game, an action video game can enhance capacity in selective attention. Although both males and females in the experimental group improved after training, the females benefited more than males (training x gender: $F(1,8) = 14.787$, $p = .005$). Their errors were greatly reduced at all eccentricities

compared to their pre-test performance. After playing with the action video game for only ten hours, the gender differences on the UFOV task were largely reduced or even eliminated. In contrast, neither males nor females in the control group showed significant improvement after playing with the non-action video game.

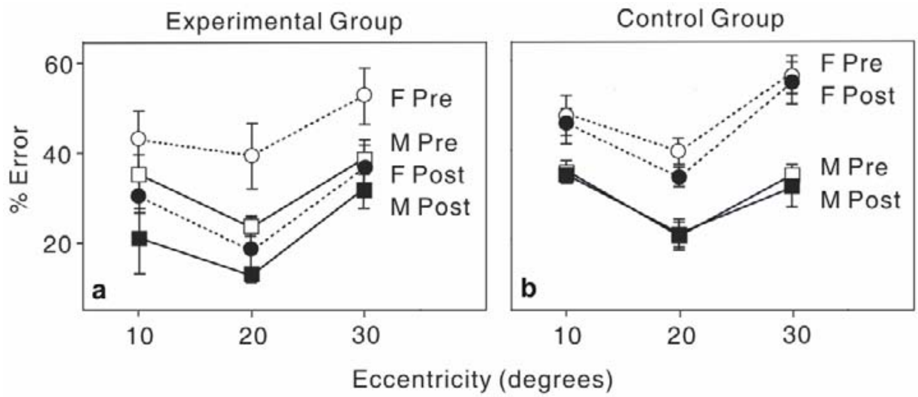


Fig. 3. Performance (percentage errors) of males and females in both groups before and after training: a) males (M) and females (F) in the experimental group before (Pre) and after (Post) training with an action video game; b) males (M) and females (F) in the control group before (Pre) and after (Post) training with a non-action video game

Experiment 2 has shown that action video gaming can improve an individual's performance on the UFOV task. Moreover, comparison with the control group confirms that this enhancement in performance is due to expanded selective attentional capacity rather than better hand-eye coordination.

The results of Experiment 2 strengthen our contention that the differences in performance in Experiment 1 are not simply the result of a selection bias. Although we cannot compare the effects of ten hours of action video game playing in the lab with weeks, months, or even years of real-world action video game playing, our findings strongly suggest that individual differences in the spatial distribution of selective attention can be modified by playing action video games as opposed to other varieties of video games.

3 General Discussion

Our first experiment demonstrated that systematic group differences in attentional capacity exist. Action video gamers generally are superior to non-gamers on a test of selective attention. Also, males perform better than females in general and this difference is particularly large among non-gamers. Moreover, these individual differences in spatial attention seem to have some impact on their academic and career choices. Students with superior selective attentional capacities are more likely to enter science and engineering fields.

We also explored a possible contributing cause to the observed individual differences. The results from the training study (Experiment 2) supported our hypothesis that action video gaming experience could modify selective attentional capacity. This is consistent with the recent results of Green and Bavelier [1]. Our study expands on their results by showing that there is a gender difference on the UFOV task whose size is much smaller among action video game players compared to non-action video game players. Furthermore, we succeeded in reducing the large gender difference among non-gamers by training with an action video game. Although ten hours of video game playing is not enough to turn a non-gamer into a gamer, the improvement in performance that we observed is very encouraging. We have shown that it is possible to modify a relevant cognitive ability (selective attention) by appropriate training and furthermore that pre-existing group differences may be reduced or eliminated.

We have confirmed that an individual's spatial attentional capacity can be improved by appropriate cognitive training using an action video game, and we have also shown that pre-existing group differences may be reduced or eliminated. Since selective attention is a critical ability in various occupations, the playing of action video games may make a positive contribution to enhancing the selective attentional capacities of the individuals who are subsequently attracted to these occupations.

Video games are appealing to young people. However, there are concerns that the violence in most action video games may be partially responsible for subsequent aggressive cognition and behavior [32]. A significant challenge for the video game industry is to produce games that retain the elements that require and develop attentional skills but that avoid the negative aspects which may encourage anti-social behaviour. New action video games designed specifically for training could yield large benefits for the information technology (IT) sector by promoting significant positive cognitive changes in young individuals. Such changes in cognitive functioning might subsequently be an important factor in helping to attract these individuals to occupations in IT. Although our study is exploratory and limited in scope, our results suggest that playing action video games may yield benefits that go far beyond mere entertainment.

Acknowledgments. This work was supported by grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada and Communications and Information Technology Ontario (CITO).

References

1. Green, C.C., Bavelier, D.: Action Video Game Modifies Visual Selective Attention. Nature 423, 534–537 (2003)
2. Ball, K.K., et al.: Age and Visual Search: Expanding the Useful Field of View. Journal of Optical Society of America 5, 2210–2219 (1988)
3. Myers, R.S., et al.: Relation of Useful Field of View and Other Screening Tests to On-Road Driving Performance. Perception and Motor Skils 91, 279–290 (2000)
4. Wickens, C.D., Hollands, J.G. (eds.): Engineering Psychology and Human Performance. Prentice Hall, Upper Saddle River New, Jersey (1999)
5. Cowan, N.: Attention and Memory: An Intergrated Framework. Oxford University Press, New York (1995)

6. Wickens, C.D., et al.: Attention-Situation Awareness (A-SA) Model of Pilot Error. A technical report for NASA Ames Research Center (2005), http://www.humanfactors.uiuc.edu/Reports&PapersPDFs/TechReport/04-15.pdf
7. Green, C.S., Bavelier, D.: Effect of Action Video Games on the Spatial Distribution of Visuospatial Attention. Journal of Experimental Psychology: Human Perception and Performance 32, 1465–1478 (2006)
8. Bandura, A. (ed.): Self-efficacy: The Exercise of Control. W. H. Freeman, New York (1997)
9. Eccles, J.S., Wigfield, A., Schiefele, U.: Motivation to Succeed. In: Eisenberg, N. (ed.) Handbook of Child Psychology, pp. 1017–1095. Wiley, New York (1998)
10. Wigfield, A., et al.: Sex Differences in Motivation, Self-Concept, Career Aspiration and Career Choice: Implications for Cognitive Development. In: McGillicuddy-De Lisi, A., De Lisi, R. (eds.) Biology, Society, and Behavior: The Development of Sex Differences in Cognition, pp. 155–181. Ablex Publishing, Westport (2002)
11. Miyake, A.: Individual Differences in Spatial Abilities. In: Shah, P. (ed.) The Cambridge Handbook of Visuospatial Thinking, pp. 121–169. Cambridge University Press, New York (2005)
12. Halpern, D.F. (ed.): Sex Differences in Cognitive Abilities. Lawrence Erlbaum Associates Press, Hillsdale, New Jersey (1992)
13. Kimura, D. (ed.): Sex and Cognition. The MIT Press, Cambridge, Massachusetts (1999)
14. Fink, A., Neubauer, A.C.: Individual Differences in Time Estimation Related to Cognitive Ability, Speed of Information Processing and Working Memory. Intelligence 33(1), 5–26 (2005)
15. Petros, T.V., Beckwith, B.E., Anderson, M.: Individual Differences in the Effects of Time of Day and Passage Difficulty on Prose Memory in Adults. British Journal of Psychology 81(1), 63 (1990)
16. West, R.L., Crook, T.H., Barron, K.L.: Everyday Memory Performance Across the Life Span: Effects of Age and Noncognitive Individual Differences. Psychology and Aging 7(1), 72–82 (1992)
17. Behrmann, M., Geng, J.J., Shomstein, S.: Parietal Cortex and Attention. Current Opinion in Neurobiology 14, 212–217 (2004)
18. Halligan, P.W., et al.: Spatial Cognition: Evidence from Visual Neglect. Trends in Cognitive Sciences 7(3), 125–133 (2003)
19. Heinke, D., Humphreys, G.W.: Attention, Spatial Representation, and Visual Neglect: Simulating Emergent Attention and Spatial Memory in the Selective Attention for Identification Model (SAIM). Psychological Review 110(1), 29–87 (2003)
20. Akhtar, N., Enns, J.T.: Relations Between Convert Orienting and Filtering in the Development of Visual Attention. Journal of Experimental Child Psychology 48, 315–334 (1989)
21. Levinoff, E.J., et al.: Selective Attention Impairments in Alzheimer's Disease: Evidence for Dissociable Components. Neuropsychology 18, 580–588 (2004)
22. Shalev, L., Tsal, Y.: The Wide Attentional Window: A Major Deficit of Children with Attention Difficulties. Journal of Learning Disabilities 36, 517–527 (2003)
23. Collaer, M.L., Hines, M.: Human Behavioral Sex Differences: A Role for Gonadal Hormones During Early Development. Psychological Bulletin 118, 55–107 (1995)
24. Hines, M. (ed.): Brain Gender. Oxford University Press, New York (2004)
25. Lohman, D.F., Nichols, P.D.: Training Spatial Abilities: Effects of Practice on Rotation and Synthesis Tasks. Learning and Individual Differences 2, 67–93 (1990)

26. Richards, E., Bennett, P.J., Sekuler, A.B.: Age Related Differences in Learning With the Useful Field of View. Vision Research 46, 4217–4231 (2006)
27. Cho, B.-H., et al.: The Effect of Virtual Reality Cognitive Training for Attention Enhancement. CyberPsychology & Behavior 5(2), 129–137 (2002)
28. McClurg, P.A., Chaillé, C.: Computer Games: Environments for Developing Spatial Cognition. Journal of Educational Computing Research 3(11), 95–111 (1987)
29. De Lisi, R., Cammarano, D.M.: Computer Experience and Gender Differences in Undergraduate Mental Rotation Performance. Computer in Human Behavior 12(3), 351–361 (1996)
30. De Lisi, R., Wolford, J.L.: Improving Children's Mental Rotation Accuracy with Computer Game Playing. The Journal of Genetic Psychology 163(3), 272–282 (2002)
31. Green, C.S., Bavelier, D.: Action Video Game Experience Alters the Spatial Resolution of Vision. Psychological Science, (in press) (2007)
32. Anderson, C.A., Bushman, B.J.: Effects of Violent Video Games on Aggressive Behavior, Aggressive Cognition, Aggressive Affect, Physiological Arousal, and Prosocial Behavior: A Meta-Analytic Review of the Scientific Literature. Psychological Science 12, 353–359 (2001)

The Effect of Traffic on Situation Awareness and Mental Workload: Simulator-Based Study

Xueqin Hao[1,2], Zhiguo Wang[1,2], Fan Yang[1,2], Ying Wang[1,2], Yanru Guo[1,2], and Kan Zhang[1]

[1] State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences,
Beijing 100101, China
[2] Graduate School of the Chinese Academy of Sciences,
Beijing 100049, China
haoxq@psych.ac.cn

Abstract. In the present study, we investigated the effects of the different traffic on the driver's situation awareness and the mental workload (MWL). The task used in this study was a medium fidelity, 3-dimensional simulation of a driving environment. The simulation required participants to drive the user's car and perform a real-world driving task. After the simulated driving, participants were asked to complete two tests which assessed their situation awareness (SA). The mental workload measures in this study consisted of the physiological measures and the subjective assessment. Every participant performed two different traffic simulated driving conditions, one was low traffic, the other was high traffic. The results showed that with the increasing of traffic, the driving performance did not worsen, however participant's mental workload increased, at the same time, the participant's situation awareness performance deteriorated. Meanwhile, our results also demonstrated that recall-based SA test and recognition-based test was heterogeneous.

Keywords: situation awareness, workload, simulated driving.

1 Introduction

The research described here investigated situation awareness (SA) in the real-time task of driving. In particular, we are interested in how drivers monitor the location of vehicles around them. Situation awareness is an important concept in applied research, originating as a description what makes up the ace factor in 1980's [1]. With the development of complex machine system, it is difficult to entirely explain how the human error results in the disastrous accidents only according to those conventional cognitive concepts. Furthermore, the assessment of interface and system design is challenging in the complex system. Thus the cognitive concept of Situation Awareness (SA) offers a new perspective. Today, it is widely accepted that good SA is required for successful and safe performance in tasks such as flying and air traffic control (ATC) [2]. In contrast, low SA has been linked to operator error [3]. The

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 288–296, 2007.

majority of situation awareness researches refer to flight and ATC. Although there are some similarities between the domains of flying and driving, there is paucity of empirical studies of SA in driving.

There are diverse definitions of situation awareness in the literature [4]. Arguably, the most accepted is by Endsley, who defines it as “The perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future [5].

As Wickens (2001) suggests, the nature of “the situation” of which awareness is maintained needs to be specified [6]. Here we will deal only with the driver’s awareness of the locations and properties of other nearby vehicles on the road way. Perhaps this should be called “traffic awareness”.

Johannsdottir et al (2003) suggested a Cognitive Systems Engineering (CSE) framework and argued that the underlying representation must be distinguished from SA [7]. In the CSE framework the theoretical concept at the centre is the Dynamic Mental Workload (DMM) which captures the constant representation and updating of the system and the task environment in working memory (WM). The framework also includes three empirical or measured constructs that reflects on the DMM, namely SA, workload, and task-relevant performance. They suggested that using converging measures of the three central concepts: SA, workload, and task-relevant performance is more adequate for assessing the complex operator-machine interaction.

The researchers who focused on the traffic situation awareness seldom were concerned for the change of mental workload simultaneously. In the present study, participants’ SA, workload, and driving performance were measured, the central objective here is that using converging measures to explore how the three constructs change with the change of the traffic.

There were two hypotheses in our study: with the increasing of the traffic, participants’ mental workload become heavier, SA becomes worse, and driving performance deteriorated. It was also hypothesized that for the same driving scene, the SA performance firstly tested was better than one tested lastly.

2 Methods

2.1 Participants

Forty-three drivers (23 skilled drivers vs. 20 novices) participated in this experiment. The skilled drivers were those who had driven for more than 3 years with more than 20,000 km, while the novice drivers were those who just got their license with little driving experience. 22 were males, and 21 were females, aged from 20 to 37. The average was 28.6 years (*SD* = 4.9).

Latin square design was implemented to counterbalance potential order effect and learning effect.

2.2 Experimental Design

There were two different experimental designs for different dependant variables. For the four SA performance indexes and the driving performance index, a 2 (the order of tests: firstly situation judging test, firstly situation resuming test) × 2 (traffic: low,

high) × 2 (gender: male, female) × 2 (experience: novice, skilled) four-factor mixed experiment design was applied. For the workload indexes, a 2 (traffic: low, high) × 2 (gender: male, female) × 2 (experience: novice, skilled) three-factor mixed experiment design was used. While the driving experience and gender were between-subjects variables, and the order of test and the traffic were within-subjects variables.

In this study, participants were asked to perform two traffic conditions. In the low traffic, there were four vehicles on the screen around the user when the moving screen disappeared. In the high traffic, there were six vehicles on the screen when simulated driving task finished.

The order of tests was assigned as follows: In every traffic condition, participants needed to perform two blocks. In one block, participants were measured firstly with the situation resuming test (SRT), then with the situation judging test (SJT). In other block, the order of test was reversed. Every participant needed to finish 4 blocks experiments, in all 24 trials.

Dependent variables included: driving performance, SA performance, physiological indexes and subjective evaluations of MWL.

2.3 Materials

The experimental task was conducted using a medium fidelity, fully configured driving simulator, which was developed ourselves to measure SA performance and driving performance. It consisted of two parts. One part included a physical steering wheel, physical gas and brake pedals. Another part involved a PC-based simulated driving and SA measuring system.

The simulation environments included a three-lane highway presented from an egocentric viewpoint inside the driver's vehicle. Participants had full use of the rear view mirror as well as both side mirrors providing them with a complete view of the traffic behind them. Surrounding traffic was located on the two lanes driving in the same direction as the participant's car. On some trials, incident would occurred that required a driving performance , for example , a car could move into the driver's lane ahead of (or behind) while moving slowly (or fast) enough that it would collide with the user car. Participants could avoid hazards to accelerate or to decelerate, or move to the lane on the left, or move to the lane on the right. Each scene was scripted to stop at a random location after 60-80 seconds of run time. The scene was updated at a rate of 30 Hz. Drivers must keep track of their route location, the position and speed of their own and other vehicles. As participants were also asked to complete two SA measures which involved the awareness of position nearby vehicles. They were situation resuming test (SRT) which was similar with Gugerty's methods and situation judging test (SJT) which was developed ourselves [4]. At the same time simulator automatically recorded a series of data which consisted of the position and speed of user's car at every second. Furthermore, the simulator automatically collected data regarding the actual location of the cars relative to the participants' vehicle at the end of each trial.

SRT: In this test, the SA performance was assessed using a post-trial question (position grid) which the participants used to indicate the position of cars relative to their own vehicle at the time each trial ended on a bird's-eye view of the road. The bird's-eye view showed the road 20 car lengths ahead of or behind the driver. The

driver's car was in red in the correct lane. Participants could use mouse and Delete buttons to move or delete cars they had placed on the road. When participants finished recalling all the car locations for a scene, they clicked on the DONE button and then continued.

SJT: In this test, participants were presented four driving scenes and were asked to judge which scene was the correct car location scene at the time each trial ended. There was only one correct situation scene in the four driving scenes. Three driving scenes as distractors were produced by changing one car's location of correct driving scene. The changing rule was to move one car from its original location to the symmetrically opposite lane. Three cars moved were respectively in the near section, middle section, and far section. When participants finished four judgments, the next trial was automatically presented by the simulator.

Simulated driving task was presented with a Dell Latitude D500 notebook.

Physiological indexes included heart rate, heart rate variability index -LF/HF, which were measured by recording equipment -- KF2 dynamic multi-parameter bio-signal detector.

Subjective evaluations of MWL were measured by NASA-LTS questionnaire, including six items for six dimensions: mental demand, physical demand, time pressure, self-evaluation, effort and frustration. Participants were asked to give a value between 0 (not at all) and 20 (extremely) to indicate their subjective feelings in the temporal manipulation on every dimension.

2.4 Procedures

The participants were tested individually in a quiet room. Each participant completed the entire experiment in one day according to the following procedures: (1) 5 min of general instruction on the experimental task; (2) about 15 min of training, including three entire trials; (3) 5 min of instruction of the SA questionnaire and subjective workload rating scale to be administrated during experimental trials; and (4) four 15-min formal experimental blocks, including the two SA tests and the subjective workload rating. They were low traffic firstly SJT, low traffic firstly SRT, high traffic firstly SJT, and high traffic firstly SRT. The presenting order of these four blocks was balanced by Latin square to participants, however two conditions in same traffic was tested in succession.

The MWL subjective workload rating scale were administrated after every traffic conditions tasks, thus each participant had to take two times of evaluation on subjective MWL. The physiological recording began before the practice and stopped after the formal experiment.

3 Results

Here, the results of the traffic and the order of SA tests were presented. The results of individual difference about gender and driving experience would be presented in other article.

The SA data included four indexes: the score of resuming direction (SXY), the estimating error (RE), the total score of judging (SJ) and the react time of judging

(RT). Collision number (SB) was recorded automatically by simulator as driving performance.

Both x- and y-coordinates for each car were collected when the simulation was frozen at the end of each trial. The y-coordinates measured the distance between each car and the participants' car, while the x- coordinates measured whether the car was positioned in the right of or the light lane. These data were compared to the participants' evaluation using the grid positioning sheet.

For SXY, participants could get two points if they put a car into the correct quadrant regardless of the distance error. If the direction of x- or y- coordinate was right, they could get one point. Then sum up all car's scores of direction and averaged it then we got the SXY. RE were calculated as the difference between the computer value and the participants' estimation of car distance. Missed cars or added cars were therefore given the value of 120 meter (maximum error). SJ was the sum of four judging scores. Participant could get one point if one judgment was correct. The correct judgment only could get 0.5 point if participants gave two "true" judgments and one was correct. If participants gave "true" judgment more than two times, the correct judgment could not get points.

In the present study, we received 33 effective data at last (expert = 16, male = 8, female = 8; novice = 17, male = 9, female = 8). All trials from 33 participants, 24 trial each, 792 trials in total, were collected.

3.1 SA Performance

Table 1 showed the results of SJ performance in different conditions. The data reported in this article were analyzed by means of a repeated measures analysis based on the multivariate general linear model. A 2 (the order of tests: firstly SJT, firstly SRT) × 2 (traffic: low, high) × 2 (gender: male, female) × 2 (experience: novice, skilled) repeated measures analysis of SJ revealed significant main effects of traffic [$F(1,29) = 5.243$, $p < .05$] and gender [$F(1,29) = 6.398$, $p < .05$]. Whereas the main effect of order of tests [$F(1,29) = 2.942$, $p > .05$] was not significant. No interaction was significant. The results showed SJ decreased along with the increase of traffic.

Table 2 showed the results of RT performance in different conditions. The four factor repeated measures analysis of RT revealed significant main effects of traffic [$F(1,29) = 6.263$, $p < .05$], the order of tests [$F(1,29) = 11.934$, $p < .01$], and gender [$F(1,29) = 4.13$, $p < .05$], whereas all interactions were not significant. The results indicated RT increased along with the increase of traffic.

Table 1. The SJ in different conditions

SJ	N	Mean	SD
SJ_DP	33	3.30	0.49
SJ_DH	33	3.17	0.47
SJ_GP	33	3.09	0.54
SJ_GH	33	2.99	0.52

Table 2. The RT in different conditions

RT	N	Mean	SD
RT_DP	33	3.61	1.42
RT_DH	33	3.50	1.31
RT_GP	33	4.42	1.62
RT_GH	33	3.74	1.05

It is interesting that RT in firstly SJT condition was longer than the firstly SRT condition regardless of traffic. This was not consistent with our hypothesis.

Table 3 showed the results of SXY in different conditions. The four factor repeated measures analysis of SXY revealed significant main effects of traffic [$F(1,29) = 23.486$, $p < .001$], and the order of tests [$F(1,29) = 12.291$, $p < .01$], whereas all interaction wasn't significant. The results indicated that the more traffic resulted in the worse SXY performance. It is also interesting that SXY in firstly SRT condition was worse than the firstly SJT condition regardless of traffic. This was not also consistent with our hypothesis.

Table 3. The SXY in different conditions

SXY	N	Mean	SD
SXY_DP	33	0.86	0.48
SXY_DH	33	0.84	0.60
SXY_GP	33	0.78	0.82
SXY_GH	33	0.73	0.98

Table 4. The estimate error in different conditions

RE	N	Mean	SD
RE_DP	33	21.59	14.93
RE_DH	33	26.55	13.98
RE_GP	32	33.71	15.70
RE_GH	33	39.81	19.16

Table 4 showed the results of RE in different conditions. The four factor repeated measures analysis of RE revealed significant main effects of traffic [$F(1,29) = 47.426$, $p < .001$], and the order of tests [$F(1,29) = 21.082$, $p < .01$]. All interactions were not significant. The results indicated that RE increased along with the increasing of traffic. It is also interesting that RE in firstly SRT condition was larger than the firstly SJT condition regardless of traffic. This was not consistent with our hypothesis.

The results of the SA performance indicated that high traffic resulted in worse SA. There were also some unexpected result for the order of tests : Except for the RT

performance, the other three SA performances were better in conditions when participants implemented firstly SJT than those in conditions when participants implemented firstly SRT.

3.2 Driving Performance

Table 5 showed the results of SB in different conditions. The four factor repeated measures analysis of SB did not reveal any significant main effects whereas there was a significant interaction of EXPERICE * GENDER [$F(1,28) = 4.607$, $p < .05$]. The results showed that SB had not been influenced by the traffic and the order of tests.

The driving performance showed that the traffic and the order of tests did not affect the driving performance in present study. Maybe this index was not enough sensitive for drivers' controlling performance.

Table 5. The collision number in different conditions

SB	N	Mean	SD
SB_DP	33	1.22	0.47
SB_DH	33	1.25	0.55
SB_GP	33	1.23	0.59
SB_GH	32	1.29	0.63

3.3 Workload Performance

Table 6 showed the results of six terms of NASA-TLX. In the workload data, we only analyses the effect of the traffic and gender and experience. The three factor(traffic, gender, and driving experience) repeated measures analysis of six dimension of NASA revealed significant main effects of traffic in three dimension which included mental required [$F(1,28) = 14.768$, $p < .001$], time pressure [$F(1,28) = 9.215$, $p < .001$], self-estimate [$F(1,28) = 6.009$, $p < .05$]. The results showed that with the increasing of traffic, the score of mental required, time pressure increased, whereas the score of self-estimate decreased.

Table 6. Subjective mental workload in different traffic conditions

Subjective mental workload	Low traffic		High traffic	
	M	SD	M	SD
MR	11.75	3.89	14.06	3.56
PR	6.156	4.0	6.69	3.6
TP	9.63	3.66	11.13	3.99
SF	9.81	3.76	8.16	3.9
EF	13.47	2.87	14.53	3.39
FR	8.97	3.98	10.25	4.04

3.4 Physiological Data

Table 7 showed the results of two physiological indicators. The three factor (traffic, gender, and driving experience) repeated measures analysis of HR revealed a significant main effects of traffic [$F(1,28) = 5.387$, $p < .05$]. The three factor repeated measures analysis of HRV revealed significant main effects of traffic [$F(1,28) = 4.872$, $p < .05$] and gender [$F(1,29) = 7.577$, $p < 0.01$]. The results showed that with the increasing of traffic, heart rate and heart rate variability increased.

Table 7. The physiological indexes in different traffic conditions

Physiological indexes	Low traffic		High traffic	
	M	SD	M	SD
Heart rate	72.22	8.28	73.43	9.81
LF/HF	2.41	1.41	2.77	1.55

4 Discussions

The results in present study indicated that with the traffic increasing, participants aroused more energy to cope with the complex task, while their subjective mental workload increased and their SA performance worsen. These results confirmed our hypothesis.

During stimulated driving, in order to track surrounding traffic, participants must shift attention to locations in the forward visual field as well as to rear- and side-view mirrors. Thus, this task requires not only perceptual tracking but also the use of dynamic spatial working memory. Therefore, when the traffic were increased, firstly drivers working memory load were added, secondly, the difficulty of avoiding hazardous also was added in that surrounding excessive vehicles diminished the safe field. In our study, two SA measures were memory-based. Thus, it is not surprising that the increased traffic resulted in worse SA performance.

But, there was a result which was not consistent with our hypothesis that participants' driving performance was not influenced significantly by the change of traffic. There were two reasons. Firstly, the reason is that we selected an index which it was not sensitive enough to the experiment manipulation. Thus, some more sensitive driving performance indexes would be necessary in the future, for example, departure distance. Secondly, the perceptual tracking subtask was not same affected by the memory workload as the memory-based SA performance.

SA performance which was measured with two methods indicated SA declined when the traffic increased. This implied that SA is a sensitive index. Furthermore, these results were consistent with Gugerty's results. But, there were some difference between the two studies. Generally, we decreased participants' memory load in SA tests. In SRT, participants were supplied the amount and color of surrounding cars, they only need put these cars on the correct locations. In SJT, participants were presented four choices, they only need to judge which one was correct. In other words, SJT was a recognition task.

We hypothesized that the SA performance of measure firstly implemented would better than those of tested lastly, but it was not the case. The results in present study were inconsistent with our hypothesis: Except for the RT in SJT, other three SA performance were better in conditions which participants were measured firstly with SJT than those conditions which participants were measured firstly with SRT. Maybe it result from the different dependence of two tests on participants' memory: In SJT, participants were presented four driving scenes, driving scenes presented in SJT were radically consistent with participants' mental model about driving task. In other words, the driving scenes would be helpful to maintain participants' mental model. While the resuming test did not provide a replaceable perception represent. So SJT would slow the vanishing of the participants' driving dynamic mental model. Thus our results also indicated that two SA tests were heterogeneous.

References

1. Kamilla, B.A.: Situation Awareness and Working Memory: an Integration of an Applied Concept with a Fundamental Cognitive Process, dissertation (2004)
2. Endsley, M.R.: Theoretical Underpinnings of Situation Awareness Analysis and Measurement. Lawrence Erlbaum Associates, Mahwah, NJ (2000)
3. Jones, D.G., Endsley, M.R.: Sources of situation awareness errors in aviation. Aviation, Space and Environmental Medicine 67, 507–512 (1996)
4. Mathews, M.L., Byrant, D.J., Webb, R.D.G.: Situation Awareness and Driving?Literature Review and Analysis (No. T8080-8-1348/001/ss). Transport Canada, Montreal, QB (1999)
5. Endsley, M.R.: Measurement of situation awareness in dynamic systems. Human Factors 37, 65–84 (1995)
6. Wickens, C.D., Hollands, J.G.: Engineering Psychology and Human Performance. Prentice Hall, New Jersey (1999)
7. Johannsdottir, K.R., LeEevre, J.-A., Herdman, C.: Cognitive system engineering (CSE) framework for evaluating cockpit interfaces. In: The Proceedings of The Twelfth International Symposium on Aviation Psychology, Colombus, OH, The Ohio State University (2003)
8. Gugerty, L.J.: Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory. Journal of Experimental Psychology Applied 3, 42–46 (1997)

Multi-window System and the Working Memory

Ayako Hashizume[1], Masaaki Kurosu[2], and Takao Kaneko[3]

[1] Graduate School of Human Sciences, Waseda University, Japan
[2] National Institute of Multimedia Education (NIME), Research and Development Division, Japan
[3] School of Human Sciences, Waseda University, Japan
hashi-aya@toki.waseda.jp

Abstract. This paper deals with the issue of the working memory load in relation to the multi-window system and explains the reason why multi-window and multi-monitor systems are better for the window operation in accordance to the structure and the function of the working memory. In the last part of this paper, a model revised from Card, Moran and Newell is proposed to explain the working memory load.

Keywords: working memory, multi-window system, user interface, dual display, memory model.

1 Introduction

Multi-window system is now widely used where multiple windows can be displayed at the same time and the currently active window will be positioned at the top. By using this system, the user can do multiple jobs on just one screen. But sometimes the window containing the necessary information will be hidden behind another one. The user will have to change the position and the size of the window accordingly so that the window containing the necessary information can be displayed better. This operation of changing window parameters requires certain time and reduces the efficiency of the operation of the user [1].

It is usually said that the overload on the working memory reduces the task efficiency. This paper will investigate the relationship between the window operation and the working memory load. Further, the paper will deal with the possible structure of the working memory and its load.

2 Window Display Patterns in the Multi-window System

Users open many windows at the same time, especially in the multi-window system so that users can use many application programs at the same time. Sometimes users leave the window where they have finished the job.

There are three patterns of window display in the multi-window system, namely, the overlapping windows, the cascade windows, and the tile view (Fig. 1).

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 297–305, 2007.

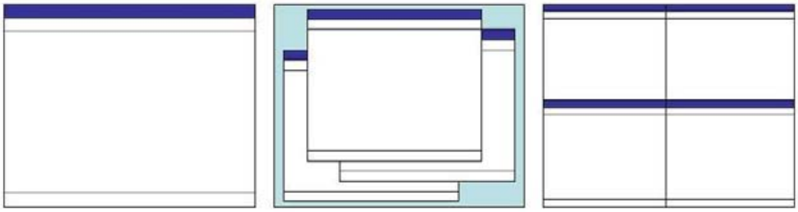


Fig. 1. Three patterns of window display: overlapping window, cascade, tile view

The overlapping windows is to display multiple windows one on the other. The cascade windows is to display windows one on the other with a slight displacement. The tile view is to juxtapose windows. This pattern was adopted in early operating systems such as Windows 1 or 3.1. At that time, the graphic power of the PC was not high so that the overlapping and the cascade patters were avoided in order to limit the use of CPU resource.

Current multi-window system frequently adopts the overlapping window. But this pattern sometimes hide the window that contains the necessary information, hence the user will have to change the location and the size of the window frequently. This operation is time absorbing and inefficient, so that the user will feel the mental stress [2].

The tile view requires less operation than the overlapping windows. So the application software sometimes adopts this pattern in order to reduce the menu operation. An example of the tile view in an application program is shown in Fig. 2.

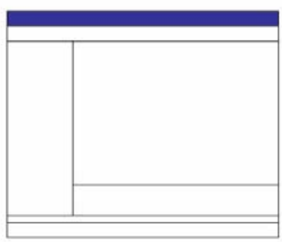


Fig. 2. Example of the tile view in an application program

Although the tile view has the merit of reducing the amount of operation, it has the deficiency of limiting the area size for an application program to display the information. This is the reason why the overlapping windows are now more prevalent than the tile view.

There are many systems proposed to solve this problem, but they tend to focus just on how to deal with the hidden windows. The optimal display arrangement depends upon the kind of applications, the task and the user preference. Thus, it is generally difficult to determine the optimal window arrangement uniquely.

Based on these considerations, it could be said that the dual monitor system will have both merits of the overlapping windows and the tile view (Fig. 3).

Fig. 3. Dual monitor system

In the dual monitor system, the user can use the same keyboard and the mouse for both monitor so that there will be less chances of doing inefficient window operations. This system is said to keep the mental workload at a certain low level and thus increases the task efficiency. In the next chapter, the relationship of this system and the efficiency of operation will be discussed in accordance to the human information processing, especially the working memory load.

3 Working Memory (WM)

The concept of working memory (WM) was originally proposed by Baddeley and Hitch (1974) in order to emphasize the functional aspect of the short term memory (STM) [3]. Until then, the STM was regarded not with the detailed processing function but just a box of simple buffering function. Baddeley and Hitch positioned the WM as to have both of the retaining function and the processing function.

It is still difficult to precisely define the WM, but it could be defined as “the memory function to process and retain the information necessary for conducting a certain task in accordance to the situational change and the progress of the task”.

WM is frequently regarded as a “box” to retain the information for a short time in order to conduct a task. But it is not a structurally independent “box” [4]. Rather, it is a set of mental mechanisms.

The role of the WM is not limited to the memory function. It is strongly related to such higher and complex cognitive functions as thinking, inference, problem solving and decision making. It is active in a complex recognition process in everyday behavior. It is the mechanism or the process to control and adjust the information related to the task [4].

WM is important in the use of the computer. It retains the incoming information, and changes and renews the contents of information accordingly to the operation and the processing. WM is used as a memory tool for achieving the goal. This kind of mechanism is necessary for operating the computer.

3.1 Long-Term Working Memory（LT-WM）

However, the actual structure of the WM is not yet completely uncovered. When there emerges the new phenomenon that the conventional memory model cannot explain, a

revision was added to the concept of WM. The concept of the long-term working memory (LT-WM) was thus proposed by Ericsson and Kintsch [5].

The reason why the concept of LT-WM was proposed was that there are some characteristics that cannot be explained by the traditional concept of WM, especially in the information processing of experts. Experts memorize information by activating abundant knowledge and skill that have accumulated in a long span of time. Ericsson and Kintsch thought that they keep the information in the long-term memory (LTM) so that it could be used instantly afterwards. According to this idea, the information that will have to be kept in the WM could only be the minimum information that can be used for the information retrieval. Thus the mass of information can be retained systematically with the minimum load on the WM itself. At the same time, such information can be accessed instantly just as the information kept in the WM in an active status. Ericsson and Kintsch called this kind of systematic retention of information as the LT-WM in order to differentiate the concept with the conventional WM or LTM.

Long term knowledge plays an important role in the use of information stored in the WM not for the experts but for various situations. Information stored in WM is just a set of fragments. The long term knowledge is necessary to actively use that information. Physiologically, it is said that the long term knowledge is stored in the hippocampus where the storage of information and the generation of retrieved information are conducted. In addition to that, the prefrontal cortex (PFC) is working for activating the retrieval key. The inter-relationship of the hippocampus and the PFC is the function of LT-WM.

3.2 Short-Term Working Memory（ST-WM）

The short term working memory (ST-WM) corresponds to the traditional concept of WM. But due to the advent of the concept of LT-WM, this term was invented for the purpose of differentiation.

Furthermore, ST-WM is conducting the encoding of information and the control of information retrieval. With these in mind, a new model of WM is proposed in Fig. 4.

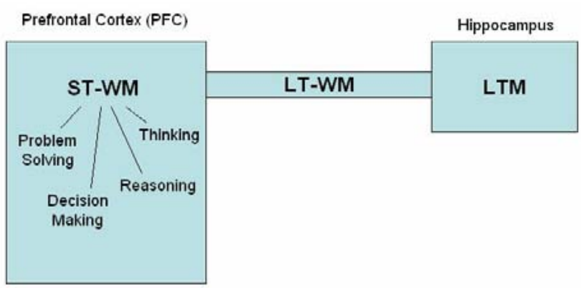


Fig. 4. A proposed model of memory including ST-WM, LT-WM and LTM

4 Load and Capacity of Working Memory

Working memory is thought to be limited in its capacity. The amount of workload varies in time depending on the task and it should be within the capacity. If it exceeds the capacity, the working memory will be overloaded and some information will be lost. And it is usually said that the task efficiently declines when too much load is given to WM. It is because there is a limitation of the processing capacity of WM. The amount of the limitation is traditionally said to be 7 plus or minus 2 chunks as Miller proposed [6]. According to this quantification of capacity limit, the amount of memory to which young people can remember is called the “chunk” and is about 7 whatever the contents to memorize is including number, letter and word. It differs for senior persons and children. Based on later researches, it was revealed that the capacity is dependent on the kinds of information, i.e. is 7 for numbers, 6 for letters, and 5 for words. Furthermore, short words requires less capacity than long words. Generally speaking, the memory capacity for information contents that can be represented as words (number, letter, and word) is related to the time to pronounce them, and is also dependent on the contextual aspects of the information content [7].

However, it is difficult to represent the capacity limitation as numbers because of various factors that may affect the working memory capacity. It is often difficult to explain the complex cognitive task if the capacity were represented as a discrete number. It can be increased when the human being is conducting a complex cognitive task. In addition to that, it was revealed that the cerebral activity relating to the working memory increases its amount by the training [8], hence it is difficult to conceive the working memory as to have a limited capacity. Hence the notion of WM to have a certain limitation is too simple to explain various phenomena. There is an idea that the processing capacity is determined based on the characteristics of the multiple processes relating to the WM. In other words, WM itself does not have the limitation but instead there is a limitation of amount of total information processing capacity reflecting various factors. Although this aspect has not yet completely specified experimentally, the idea of capacity limit of WM will be one of the crucial issues in the cognitive psychology.

In the following section, a model is presented to explain how the WM is related to the task execution in the use of computer.

4.1 The Model Proposed by Card, Moran and Newell

Card, Moran and Newell (1983) proposed a model to explain the relationship between the task performance the working memory load [9]. Fig. 5 shows the idea of this model.

This figure shows the model describing the situation where the user is entering the text by looking at the manuscript. As the progress of the task, the user conducts detailed operations including the gaze at the screen, the character typing, and the shift of the pointer. Sometimes there is a duplication of the operation such as the shift of pointer while gazing at the screen. This kind of detailed duplicated operation will have to be stored in the stack (push) and then be taken out of the stack (pop). Such stack operations can be the load to the WM.

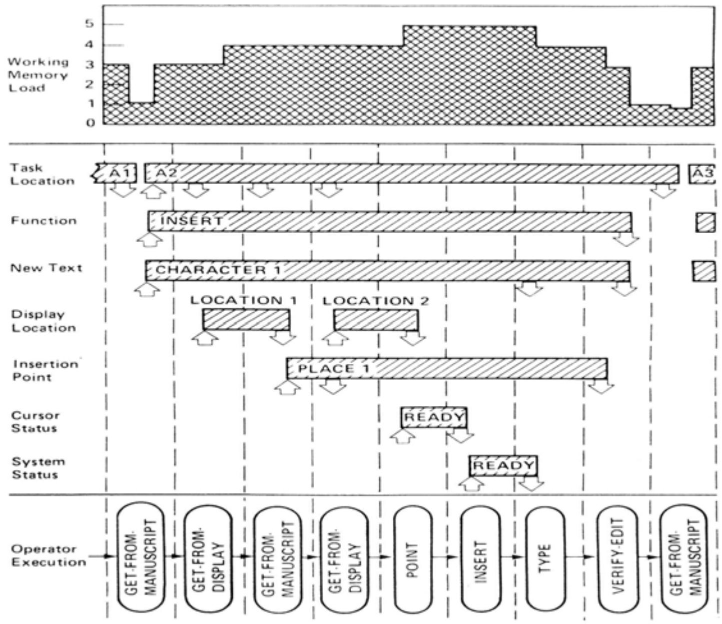


Fig. 5. The model proposed by Card, Moran and Newell to explain the working memory load during the execution of a task (Card, Moran and Newell, 1983)

This original model of Card, Moran and Newell shows how the memory load will change as the operation progresses. But it has a deficiency that should be reconsidered: the physical workload such as the gaze on the screen and the mouse operation are categorized as the working memory load. The load for the operator during some task is not only the working memory load as is shown in Fig. 5. It is better to treat the physical workload and the mental workload differently, considering the fact that the working memory load can be generated in various kinds of factors. Based on this consideration, a revised model will be proposed in the next section.

4.2 A Revised Model

The model mentioned above was originally proposed in 1983 and the situation of the computer use is somewhat different from that of today. Users today do not frequently use the paper-based manuscript to enter the text into the computer. Instead they tend to use the web site to get information while creating a new document or to use the data file stored in the disk. In other words, the degree of use of the display is much higher today than the time they proposed the model.

As such, it is more frequent that the users are using multiple windows at the same time. Fig. 6 shows such situation by applying a revised model of Card, Moran and Newell. This model splits the load to the "mental workload" and the "physical workload".

The basic idea of this revised model is to integrate the model of Fig.5 with the idea of Keystroke Level Model (KLM) that they also proposed. On the left hand side of Fig. 6, procedural steps of searching a word in a dictionary while browsing a web site

are described at the level of keystroke. It often occurs that the users will have to check the dictionary for the word they encounter in the website, especially in foreign language.

The typical sequence will be as follows and KLM operators and the time estimates can be described for each of them. The dictionary search operation will be triggered while browsing the website (Get some information on the browser (nM) 1.35*M).

(1) The user finds a new word (M = 1.35 sec)
(2) The user will have to remember the word only if he is using the single monitor (M = 1.35 sec)
(3) The user opens or activates the dictionary window (H+P+K = 0.4+1.1+0.2 sec)
(4) The user types in the word with n letters (nK+M+K) 0.2*(n+1)+1.35 sec)
(5) The user get the meaning (M = 1.35 sec)
(6) The user will have to remember the meaning of the word only if he is using the single monitor (M = 1.35 sec)
(7) The user closes or inactivates the dictionary window (P+K = 1.1+0.2 sec)
(8) The user finally understands the information on the browser (M = 1.35 sec)

And the user will iterate this process while browsing through the website.

On the right hand of Fig. 6, accumulated physical workload and accumulated mental workload are shown. Physical workload is necessary only when the user moves their hand. But the mental workload is stacked with the consciousness that the user is browsing the website as the bottom level. Upon this fundamental level, the load for finding the word, remembering it, typing it, getting its meaning and remembering it, then finally understanding the meaning of information on the browser take certain amount of mental workload. What is important here is the mental workload for remembering the word and remembering its meaning are necessary processes only for the single monitor system as can be seen in Fig. 7-a,b,c.

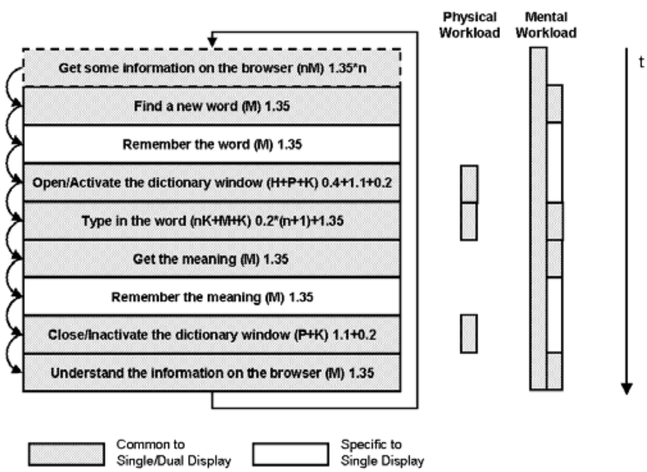


Fig. 6. Revised Model

Fig. 7-a,b. Dictionary search while browsing the website on a single monitor system

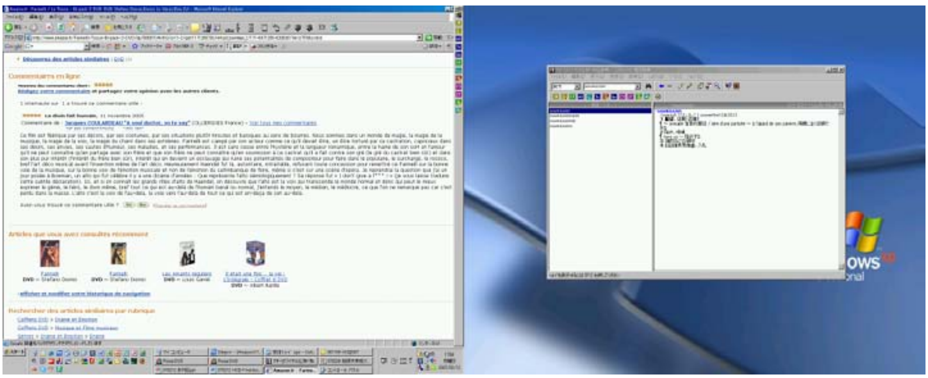


Fig. 7-c. Dictionary search while browsing the website on a dual monitor system

5 Conclusion

Based on the concept of working memory, the use of multi-window system was analyzed. A new use of the window system is now possible using dual monitors by putting each window on each side of monitors. Authors proposed a revised model based on the working memory model and the Keystroke Level Model originally proposed by Card et al. and analyzed the difference between them, thus put the prediction that less mental workload will be necessary for the dual monitor system.

Although our study revealed the difference of mental workload based on the concept of working memory, there are some points that require the further research.

(1) In Fig.5 and Fig. 6, it was avoided to decompose the load of each operation in detail. The load is represented as having the same amount at any time for any kind of operation, although the length of time to be loaded may vary. But in the real situation, the amount of the load must be different depending on the kind of operation. Furthermore, the mental workload is not the accumulation of simple loads but may be able to be decomposed more in detail because it is closely related to various factors including the working memory load. They may vary not only by the kinds of operation but also by the situation and contents of operation as well as the operators themselves. As Ericsson and Kintsch [5] proposed the concept of LT-WM, the amount of load may be

decreased in accordance with the expertise. Another possibility is that the training will make working memory more efficient as Klingberg et al. [8] proposed. In this sense, the working memory load should be considered in relation to the knowledge level of operator considering the practice and the expertise. This point should be considered further in the future research.

(2) Our approach was based on the hypothetical model and does not have the empirical evidence. The KLM is sometimes said to be less predictable due to the use of fixed parameters. This kind of model will be effective (only) when there is no real prototype. Because we already have the dual monitor system, it is necessary that we should conduct an empirical study although it is not easy to conduct the experiment in natural settings.

References

1. Tsuhara, S.: Automatic window rearrangement using fuzzy rules. In: Proceedings of ICUPC'95, pp. 682–686 (1995)
2. Arisawa, M.: Human Interface [In Japanese]. Jikkyo Shuppan Co, Tokyo Japan (1995)
3. Baddeley, A. D., Hitch, G. J.: Working memory. In: G.H. Bower (ed.), The Psychology of Learning and Motivation. Advances in Research and Theory, Vol. 8, Academic Press, 47-90 New York (1974)
4. Miyake, A., Shah, P.: Models of working memory. In: Miyake, A., Shah, P. (eds.) Toward Unified Theories of Working memory, pp. 442–481. Cambridge University Press, New York (1999)
5. Ericsson, K.A., Kintsch, W.: Long-term working memory. Psychological Review 102, 211–245 (1995)
6. Miller, G.A.: The magical number seven, plus or minus two – some limits on our capacity for processing information. Psychological Review 63(1), 81–97 (1956)
7. Hulme, C., Roodenrys, S., Brown, G., Mercer, R.: The role of long-term memory mechanisms in memory span. British Journal of Psychology 86, 527–536 (1995)
8. Klingberg, T., Forssberg, H., Westerberg, H.: Training of working memory in children with ADHD. Journal of Clinical & Experimental Neuropsychology 24, 781–791 (2002)
9. Card, S.K., Moran, T.P., Newell, A.: The Psychology of Human-Computer Interaction, pp. 392–397. Lawrence Erlbaum Associates, Hillsdale NJ (1974)

Human Performance Model for Combined Steering-Targeting Tasks

Seung-Kweon Hong[1] and Seungwan Ryu[2,*]

[1] Department of Industrial & Management Engineering, Chungju National University, 123 Geomdan-ri, Chungju, Chungchungbuk-do, Korea
[2] Department of Information System, Chung-Ang University, 72-1 Nae-ri, Anseong, Gyeonggi-do, Korea
skhong@chungju.ac.kr,rush2384@cau.ackr

Abstract. The combined steering-targeting tasks are frequently encountered within the window-type environment. For example, within a menu driven interface, the user is required to steer down a menu and then to click on a target. In this paper, human performance for these tasks was measured and a mathematical model was proposed to describe the human performance. The cursor movement in the combined steering-targeting tasks had a positive acceleration form at the starting period until about 10mm and a uniform velocity was maintained during the intermediate period and a negative acceleration was observed at the ending period. The proposed model consisted of two terms in which the first being the classical Fitts' term and the second being the steering law suggested by Drury. This model provided a good fit to the data obtained from the experiments (r^2 = 0.936) as well as the data obtained by Kvalseth (r^2 = 0.982).

1 Introduction

The human-computer interaction (HCI) community is continuously searching for useful models to either describe or predict an interaction of human behavior with computer systems. Designers use these models not only for improvement of input devices but also for evaluation of the usability of software. This paper reviews the models which describe the human behavior for the mouse control tasks and explores a new model to describe human behavior for the combined steering-targeting tasks.

The computer work using a mouse can be categorized into three types. The first is a targeting task including a point-click operation and a drag-drop operation. This task consists of (1) moving a cursor toward a target from a certain position and (2) placing a cursor on the target. Fitts' law [1] has been widely applied to this type of task. The second is a steering task, moving a cursor through a particular space with certain constraint. All of the examples include steering down both menu and its submenus, tracing a shape in a drawing program, and moving a scroll bar down a word processor or web page. A human performance model for the steering tasks is defined as a steering

* Corresponding author.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 306–315, 2007.

law [2, 3]. The third is a combined steering-targeting task. This task is also frequently encountered within the windows-type environment. For example, within a menu driven interface, the user is required to steer down a menu and then click on its target.

There may be more in the combined steering-targeting tasks compared to those in pure targeting tasks or pure steering tasks. Even though the two models were proposed in the previous studies for the tasks [4, 5], they were not concretely validated.

In this paper, both assumptions and limitations of the models which are associated with targeting tasks were reviewed first and then the review of steering tasks was conducted. A new model for steering-targeting tasks was proposed on the basis of the characteristics of these models. The proposed model was evaluated by reanalyzing the previous data and analyzing the data which obtained from new experiments. In order to find out the differences between the combined steering-Fitts' movement and Fitts' movement or steering movements, the change of cursor's moving speed during the tasks was also investigated in this study.

2 Fitts' Law and Steering Law

Fitt's law is one of the human psychomotor performance models which has been widely used in human computer interaction and it represent the predicted mean time it take to point at a target of width W over distance A:

$$MT = a + b\log_2\left(\frac{2A}{W}\right) \tag{1}$$

where a and b are regression coefficients. The logarithmic term in Equation 1 is referred to as the Index of Difficulty (ID), showing how difficult the targeting task is. The original Fitts' law was proposed in 1954 and many studies have been carried out to validate Fitts' model since then. Many alternative ID measures were also proposed [6, 7]. Although several alternatives of ID measure were suggested, the common points were the fact that the movement time is logarithmically proportional to the width (W) of the target and is inversely logarithmically proportional to the amplitude (A) of the movement. Another characteristic of Fitts' model was that it does not either describe the height of target or the trajectory of the movement.

Fitts' equation was applicable only to the movements which are visually controlled by the step for approaching to the target. Crossman [4] proposed that Fitts' equation is found to be only valid for visually controlled movement with IDs greater than the above three. If the value is lower than three, then the movements are ballistic and a different relationship should be applied to model the movement time [8].

Human movement with other type of visual feedback has been described as a steering law [2, 3]. For a cursor movement within a path with lateral constraints, human operators have to receive visual feedback to maintain a cursor within the path. Originally, such a movement was modeled for the automobile control to maintain an automobile within a lane [9, 10]. The relationship between (a) the average speed of movements along straight and circular lines and (b) the tolerance widths of the lines is shown in the following model.

$$V = kW \tag{2}$$

where k is a constant representing the controllability of the operator/vehicle system, V is a chosen velocity and W is a width of the lane.

By applying this model to the human movement, Drury [2] suggested a new model which represents a relationship between the width of the lateral constraints and the distance of the movement in self-spaced movement.

$$MT = a + b(\frac{A}{W}) \tag{3}$$

The curvilinear abscissa was introduced as the integral variable for extending this model to more generic formula: If C is a curved path then the index of difficulty for steering through this path is defined as the sum along the curve of the elementary indexes of difficulty [3].

$$MT = a + b\int_{c} \frac{ds}{W(s)} \tag{4}$$

where the integral variable s is the curvilinear abscissa and W(s) is the tunnel width at s.

From the point of movement speed’s view, Fitts' movement is different from the steering movement. The steering movement does not approach to the target and the movement speed is not changed by the whole process of the movement. A previous study showed that when velocity is unstable, the steering law is not able to provide a good prediction [11]. However, Fitts' movement has a positive acceleration and a negative acceleration in the starting period and the ending period respectively.

The movement speed in the combined targeting-steering tasks has never been examined until now. However it is possible that the combined steering-targeting tasks contain properties of both Fitts' movement and steering movement. In the starting period, a cursor movement would follow Fitts' movement by increasing the movement speed. During the intermediate period, movement speed would not increase above a certain level, because of the adjustment for maintaining a cursor within the path constraints. As the cursor approaches to the target, its movement would be decreased by the Fitts' movement. If the path was wide enough then the whole movement would not be affected by the property of steering. On the contrary, the effect of steering movement would be stronger if the path was narrow. Figure 1 presents the concept of movement speed in the combined targeting-steering task.

The lateral constraint in the Fitts' movement was started to be considered with the height of the target. In the usual Fitts' type of experiment, the target is generally much longer, in a direction perpendicular to the motion, than in the direction of the motion (e.g. the width of the target > the height of the target). If the height of the target is limited, there would be little effect of lateral constraint on the path which may be taken in moving from one target to the other.

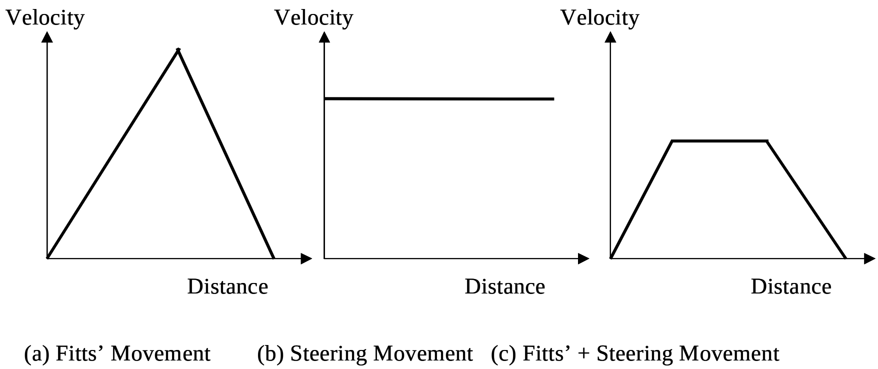


Fig. 1. Movement speeds in Fitts' movement, Steering movement and Steering + Targeting movement

The first experimental study of the effects of lateral constraint showed that of Crossman [4] in which the height of the target was varied from Wt/l6 to Wt. Crossman states (p85) "...the effect of the depth seems likely to operate according to the same law as the width." He correlated his data, from two subjects, with an Index of Difficulty, and suggested that an appropriate expression would be described as:

$$MT = a + b\log_2(\frac{A}{W_t}) + c\log_2(\frac{A}{W_p}) \tag{5}$$

where the coefficients a and b were empirically determined.

Kvalseth [5] also suggested a multiple forms of Fitts' law and applied them to the three different types of tasks: (a) similar to those of Crossman [4] with restricted target height (b) movement to a target with a center constraint (c) movement to a target via a path with lateral constraint along its entire length (straight, sinusoidal or random paths)

$$MT = a + b\log_2(\frac{A}{W_t}) + c\log_2(\frac{1}{W_p}) \tag{6}$$

In this paper, we proposed another model for the combined targeting-steering task. This model was developed by integrating Fitts' law with the steering law of Drury [2].

$$MT = a + b\log_2(\frac{A}{W_t}) + c(\frac{A}{W_p}) \tag{7}$$

In the following, the data from Kvalseth [5] were reanalyzed in terms of three models for the steering-targeting tasks. New experiments were conducted by following the reanalysis of the previous data which are aimed at investigating the effects of both tights and loosed path constraints on the movement times.

3 Reanalysis of Previous Data

3.1 The Effect of Target Height

An analysis has been made on the data obtained from the Kvalseth's experiment [5] that the height of the target ranged from 0.32cm to 2.54cm. The data has been regressed into three forms, first according to the Crossman's equation (5), secondly by following the kvalseth's equation (6), and finally by the model (7) suggested in this paper. Results were (movement time in 10^{-2} s) described as follows:

$$MT = -56.7 + 18.5\log_2(\frac{A}{W_t}) + 14.9\log_2(\frac{A}{W_p}); \quad r^2 = 0.782 \tag{8}$$

$$MT = -61.2 + 33.4\log_2(\frac{A}{W_t}) + 29.8\log_2(\frac{1}{W_p}); \quad r^2 = 0.821 \tag{9}$$

$$MT = -19.1 + 18.2\log_2(\frac{A}{W_t}) + 0.928(\frac{A}{W_p}); \quad r^2 = 0.954 \tag{10}$$

There was no significant difference in the variances indicated by the three regressions. However, the last equation (10) showed a slightly better fit.

3.2 Targets Connected by a Straight Path

Kvalseth reported the movement times in the 3 types of path (straight, sinusoidal and random paths). However, this study only concentrated on the straight path. The data for straight path were reanalyzed in terms of the three models described above. The following regressions were obtained for these data (movement time in 10^{-2} s). The model proposed in this paper provided a slightly better fit for the steering-targeting tasks.

$$MT = -196 + 24.3\log_2(\frac{A}{W_t}) + 50.3\log_2(\frac{A}{W_p}); \quad r^2 = 0.671 \tag{11}$$

$$MT = -211 + 704.6\log_2(\frac{A}{W_t}) + 101\log_2(\frac{1}{W_p}); \quad r^2 = 0.664 \tag{12}$$

$$MT = -76.5 + 24.1\log_2(\frac{A}{W_t}) + 3.29(\frac{A}{W_p}); \quad r^2 = 0.982 \tag{13}$$

4 Experiment

4.1 Apparatus and Participants

The experiment was conducted on a notebook running Windows XP with a 17-inch LCD monitor. An optical two-button mouse with a scrolling wheel was used in the

study. Software was developed in Visual Basic version 6.0. The display which is shown in Figure 2 was presented to participants in the experiment. Twelve volunteers (10 male, 2 female) participated in the experiment. Their ages ranged from 22 to 26 years old. All of them were right-handed. They were all experienced computer users, using a mouse on a daily basis.

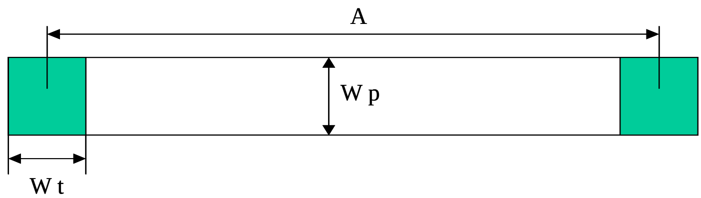


Fig. 2. An experimental apparatus for measurement of the cursor movement times

4.2 Experimental Design

The study was carried out to investigate the time to be taken for a subject steering the cursor to a target and clicking on its target. A trial was started by clicking on the left box. The subjects moved the cursor to the right box as soon as possible, without touching the boundaries of the path. The subjects clicked the box when the cursor was reached to the right box. After that, the subjects moved the cursor towards the left box and clicked it. This reciprocal tapping task was repeated 20 times for each subject. A training session of 30 min was given to each subject before starting the main tasks.

A within-subject design was used in this study. The experimental conditions, which are given in Table 1, were applied in this experiment. The difficulty of index (Fitts' ID) in the experimental conditions ranged from about 2.5 to 7, because the tasks of very small IDs may not follow Fitts' law. The height of the target was automatically determined by the width of the path. The order of each task was randomized and each subject performed the total of 48 tasks with 20 repeated trials.

Table 1. Experimental conditions

Variables	Levels
Width of target (W_t)	4mm, 8mm, 12mm
Width of path (W_p)	4mm, 8mm, 12mm, ∞
Length of path (A)	30mm, 60mm, 90mm, 120mm

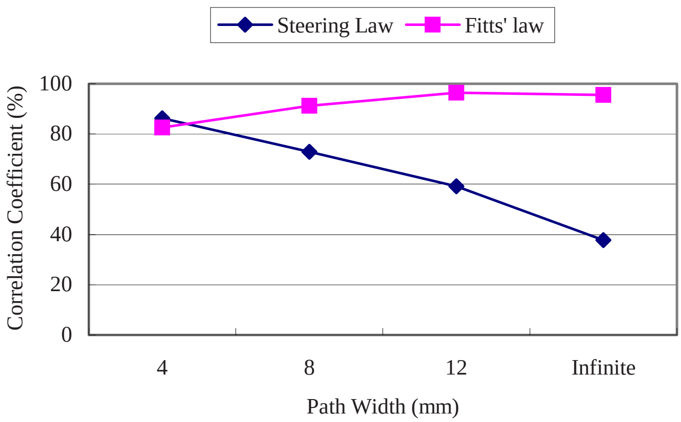
5 Analysis and Results

5.1 Mean Movement Times Under the Experimental Conditions

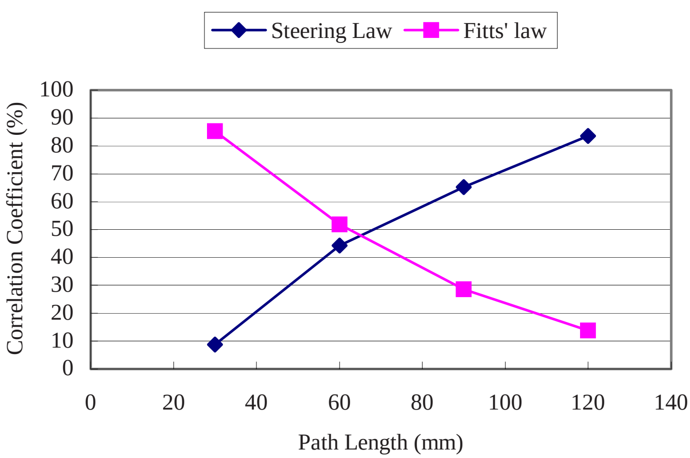
Mean movement time under each experimental condition was determined by averaging 20 movement times over 12 participants. An ANOVA was performed on the mean movement times. The mean movement times were significantly different according to the width of the target (Wt), the width of the path (Wp) and the distance of movement (A).

A significant interaction effect was observed between Wp and A. That is, if the distance of movement was long enough, the movement time would be affected by the width of path. However, if the distance of movement was short, then the movement time would not be affected by the width of the path.

If the width of the path was wide enough so that an operator did not require a visual feedback on the lateral constraints, the movement times could be described only by Fitts' law. Figure 3 shows how either Fitts' law or Steering law can describe the mean movement time according to the change in both width and length of the path. Not surprisingly, the narrower the width of the path was the better fit the steering law gave and the worse fit the Fitts' law provided. The longer the length of the path was the better fit the steering law gave and the worse fit the Fitts' law gave.



(a) Change in the path width



(b) Change in the path length

Fig. 3. Correlation coefficients representing the relationship between Fitts' law and Steering law and the measured data.

The times for cursor movements were measured at each 10 mm intervals from the starting point to the ending point. Figure 4 is an example of the measurement of cursor movement times at the condition of Wt = 8mm and A = 120mm. Regardless of the path width, a positive acceleration and negative acceleration occurred at the 10mm intervals of starting and ending point. Constant velocity occurred during the intermediate period and the velocity was changed according to Wp. As shown in both Figure3 (a) and Figure 4, the effect of path width was not observed at the approximate 16mm.

On the other hand, although the effect of path width did not exist (Wp = ∞), the movement speed was changed as shown in Figure1(c). Therefore, a typical Fitts' movement was changed in accordance with either Figure1(a) or Figure1(c).

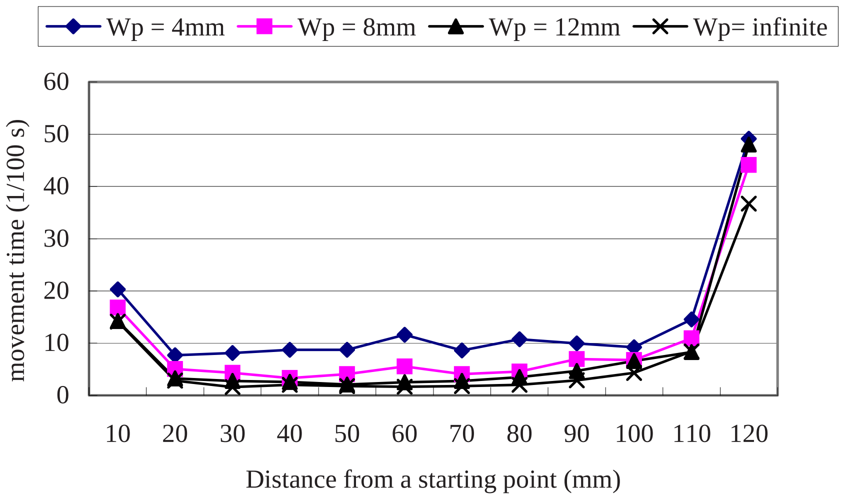


Fig. 4. Cursor movement times in the 10mm intervals at the condition of Wt = 8mm and A = 120mm

5.2 Model Fit

The values of mean movement time obtained from the experiment were fitted into the three models which were previously suggested for the steering-targeting tasks. The equations (14), (15), and (16) are the results of regression analysis. The model proposed in this study showed better fit compared to the others (r^2 =0.936 > r^2 = 0.778 > r^2 = 0.767).

$$MT = 0.029 + 0.140\log_2(\frac{A}{W_t}) + 0.151\log_2(\frac{A}{W_p}); \quad r^2 = 0.767 \tag{14}$$

$$MT = 0.757 + 0.225\log_2(\frac{A}{W_t}) + 0.178\log_2(\frac{1}{W_p}); \quad r^2 = 0.778 \tag{15}$$

$$MT = 0.197 + 0.136\log_2(\frac{A}{W_t}) + 0.028(\frac{A}{W_p}); \quad r^2 = 0.936 \tag{16}$$

6 Conclusions

This study demonstrated that the simple two-component models can be used to describe the data for movements in which there is constraint in a direction transverse to the direction of movement. The best fit model consists of two terms. The first term is the classical Fitts' term and the second is the one suggested by Drury, which is being linear in the ratio of amplitude to lateral tolerance. This model provided a good fit to the data obtained by the new experiments ($r^2 = 0.936$) as well as the data provided by Kvalseth ($r^2 = 0.982$).

The proposed model also provided a good fit to the tasks with the limited height of the target ($r^2 = 0.954$). Crossman model ($r^2 = 0.782$) and Kvalseth model ($r^2 = 0.821$) were able to make relatively good predictions, but these models were unable to provide good fits in the tasks with path constraints ($r^2 = 0.671$, $r^2 = 0.664$).

The lateral constraints of path became stronger when the width and the length of path is narrower and longer relatively. The cursor moving speed in the typical combined steering-targeting tasks are shown in Figure 1 (c). However, if the moving distance (A) was short then the movement speed might occur as those shown in Figure 1(a). In the future study, the proposed model can become more generalized for the more diverse forms of movement tasks with an application of a generalized steering model of Accot [3].

Acknowledgements. This work was supported in part by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD)"(KRF-2005-003-D00462).

References

1. Fitts, P.M.: The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. Journal of Experimental Psychology 47, 381–391 (1954)
2. Drury, C.G.: Movement with Lateral Constraints. Ergonomics 14, 293–305 (1971)
3. Accot, J., Zhai, S.: Beyond Fitts' law: Models for Trajectory-Based HCI Tasks, In: Proceedings of CHI '97, pp. 295–302 (1997)
4. Crossman, E.R.F.W.: The Measurement of Perceptual Load in Manual Operations. Unpublished Ph.D. Thesis University of Birmingham (1956)
5. Kvalseth, T.O.: A Generalized Model of Temporal Motor Control Subject to Movement Constraints. Ergonomics 20, 41–50 (1977)
6. Welford, A.T.: Fundamentals of Skill. Methuen, London (1968)
7. MacKenzie, I.S.: Fitts' Law as a Researh and Design Tool in Human Computer Interaction. Human Computer Interaction 7, 91–139 (1992)
8. Hoffman, E.R.: An Ergonomic Approach to Predetermined Motion Time Systems. In: Proceedings from the 9th National Conference, Institute of Industrial Engineers, Austrailia (1981)

9. Rashevsky, N.: Some Remarks on the Mathematical Aspects of Automobile Driving. Bulletin of Mathematical Biophysics 21, 299–308 (1959)
10. Rashevsky, N.: Man-Machine Interaction in Automobile Driving. Traffic Safety 65, 101–107 (1965)
11. Montazer, M.A., Drury, C.G.: A Test of the Beggs' Model for Self-Spaced Movements. Ergonomics 32, 497–511 (1989)

A Mental Workload Predicator Model for the Design of Pre Alarm Systems

Sheue-Ling Hwang[1], Yi-Jan Yau[1], Yu-Ting Lin[1], Jun Hao Chen[1], Tsun-Hung Huang[1,3], Tzu-Chung Yenn[2], and Chong-Cheng Hsu[2]

[1] Institute of Industrial Engineering & Engineering Management
National Tsing Hua University, Hsinchu,
Taiwan, R.O.C. 300
[2] Institute of Nuclear Energy Research, Lungtan,
Taiwan, R.O.C. 325
[3] Department of Industrial Engineering and Management,
National Chin-Yi Institute of Technology, Taichung,
Taiwan, R.O.C. 411
slhwang@ie.nthu.edu.tw, d927814@oz.nthu.edu.tw,
d943847@oz.nthu.edu.tw, g943848@oz.nthu.edu.tw,
toby.hwang@gmail.com, tcyenn@iner.gov.tw,
worm@iner.gov.tw

Abstract. This study investigated the operator's mental workload of the fourth Nuclear Power Plant in Taiwan. An experiment including the primary and secondary tasks was designed to simulate the reactor shutdown procedure of the Nuclear Power Plant. The performance of secondary task, the subjective mental workload and seven physiological signals of participant were measured. The Group Method of Data Handling (GMDH) was applied to integrate these physiological signals to develop a mental workload predictive model. The relationship between subject mental workload and the performance of secondary task is highly correlated with Pearson correlation coefficient as 0.691. The validity of the proposed model is very high with R2=0.85. The proposed model is expected to provide supervisor a reference value of operator's performance by giving physiological signals. Besides nuclear power plant, the proposed model could be applied to other fields such as aviation, air transportation control, driving and radar vigilance, etc.

Keywords: Mental workload; Physiological signal; GMDH; Predictor.

1 Introduction

Modern complexity system could bring heavy mental workload to their operators. The high rate of information flow, complexity of the information, numerous hard decisions and task time stress could overwhelm the system operators. On the other hand, high level of automation could lead to low mental workload [1, 2]. For most of the operators, their performance would be degraded when their mental workload is either too high or too low. Only in an appropriate level of mental workload, the

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 316–323, 2007.

operators could perform as anticipated [3, 4]. Although both high and low mental workloads would cause operator's performance degradation, most of the previous studies paid much attention on the problems caused by high mental workload rather than by low mental workload. Hence, developing an early warning system to measure the operator's mental workload is a very important issue.

Performance measurement, subjective ratings and physiological measurement have been considered as the three most general mental workload measurements [5, 6]. Among these measurement technologies, physiological measurement could record the continuous data and be less intrusive on work activities and high sensitive to the cognitive requirements of a complex task. Hence, it is more suitable for measuring real time mental workload than others [7, 8]. However, Chen and Vertegaal [9] pointed out that there has been little use of real-time physiological measure to dynamically manage the operator's mental workload during system operation.

The physiological responses to mental tasks were different for each person and the physiological response patterns were also different from task to task [8, 10]. Thus, it is important to consider the situations while one try to develop a real time mental workload predictive model. One approach to solving these problems is to record several physiological indexes and integrate them into one synthesized index by individual difference, and analyze it immediately [9, 7].

Numerous researches have successfully used EEG signals to classify the operator's mental workload [2, 9, 11]. Their results indicated that the average rate of successful classification was over 80%. However, Farmer and Brownson [5] indicated that using EEG to classify mental workload was not practical since the collecting data was difficult to be analyzed and had high noise-to-signal ratio. Furthermore, the equipment required to calibrate to each individual and be operated by trained person. Therefore, it is not suitable for using in the real field.

The purpose of the present research is to develop a real-time, non-intrusive mental workload predicative model by using Group Method of Data Handling (GMDH) to integrate seven physiological indexes into a synthesized index. Comparing this index with a performance level set by supervisor, if the value of the synthesized index was lower than the setting performance level, the early warning system would be started to alert the operators to adjust their workload. The physiological indexes used in this study were parasympathetic/sympathetic ratio (LF/HF) (X1), heart rate variability (HRV) (X2), heart rate (X3), diastolic pressure (X4), systolic pressure (X5), eye blink frequency (X6), and blink duration (X7).

2 Method

2.1 Subjects

Fifteen paid NTHU graduate students participated in this experiment. All of them had normal eyesight and good health. The mean age of the participants was 24 years.

2.2 Apparatus

Meditech ABPM-04 ambulatory blood pressure measuring device was used to measure the heart rate and blood pressure (Systolic Pressure and Diastolic Pressure).

ANSWatch TS0411 was used to measure Heart Rate Variability (HRV) and Parasympathetic/Sympathetic Ratio (LF/HF). Face lab version 4.0 was used to measure the blink frequency and blink duration.

2.3 Procedure

Prior to the experiment, the subject took a 15-minute rest, and then wore the measurement apparatus and proceeded with Face Lab adjustment. The initial physiological indexes were taken as a base line before the experiment. After the adjustment and measurement, the experimenter illustrated the experimental task and instructed the participant how to operate the system. The participant took about 30 minutes to practice the control procedure and familiarize with location of information display. There were three phases simulated the mental workload of secondary task from heavy, median to low. The subject simultaneously executed the primary tasks and the secondary tasks. The primary task was to shut down the reactor following the procedure of “shut down cooling model” provided by the Institute of Nuclear Energy Research (INER) (see Fig. 1). The shut down procedure was to decrease the power by inserting the nuclear rods or adjusting the core flow rate by turning off the pump. The secondary tasks were to complete a series of mathematical comparisons at each phase. The different mental workload levels of the secondary task were designed by different mathematical calculation. The tasks were presented in four monitors except the Face-Lab monitor interface, and their layout was showed in Fig. 2 The physiological indexes (LF/HF, HRV, heart rate, diastolic/ systolic pressure, blink frequency and blink duration) were measured prior to experiment and during each phase, and the NASA-TLX questionnaire was conducted after each phase. Each phase lasted for about thirty minutes. The seven physiological indexes were transferred into individual difference and acted as input variables for the model development. The performance of secondary task was evaluated by the error rate of the number comparison. The subjective mental workload was rated by the NASA-TLX score after each phase.

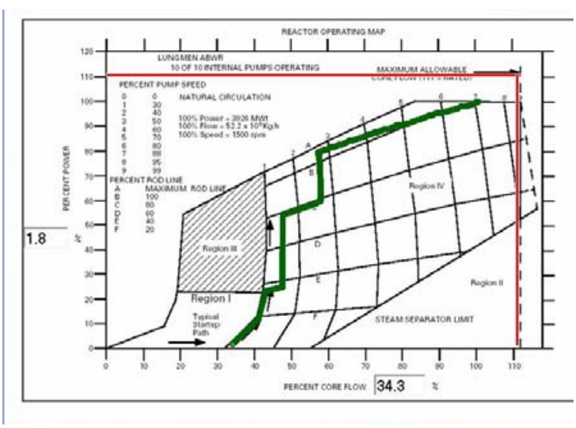


Fig. 1. The power and flow monitor interface

Fig. 2. The control interface screen layout

2.4 Analysis

The error rate of the series of mathematical comparisons at each phase was collected to compare with the corresponding scores of NASA-TLX questionnaire. The higher score of NASA-TLX questionnaire reflected higher level of mental workload during experiment. Correlation between the performance of secondary task (error rate) and the evaluation of subjective mental workload was analyzed by the Statistical Products and Services Solution (SPSS).

Group Method and Data Handling (GMDH), one of the well-know neural network methodologies, was applied to develop the predicative model. GMDH was developed in 1971 [12] and is one of the exact prediction methods in last decade. The GMDH algorithm has been widely used in various fields such as education [13], business, and vehicle factory [14]. This study investigated the relationship between seven physiological indexes and mental workload. These physiological indexes (X1~X7) were collected as input variables and then created a model to predict different degree of mental workload.

3 Results

All participants finished their primary task before time limitation and met the requirements set by INER; hence, differences in mental workload would be reflected by the performance of the secondary task.

3.1 Correlation Between Secondary Task Performance and Subjective Mental Workload

The relationship between the subjective mental workload and performance of secondary (error rate) is shown in Fig. 3. In general, the directions of two trend lines of error rate and subjective mental workload were definitely the same. Almost all subjects made fewer errors and rated lower score of NASA-TLX questionnaire as the secondary task becomes less complex.

The analysis of Pearson-product moment correlation was used to examine the relationship between secondary task performance and NASA-TLX subjective mental

workload as shown in Table 1. It indicated that the error rate and the subjective mental workload were positively correlated with each other. The correlation coefficient of 0.691 was found to be statistically significant at $p < 0.01$.

As a result, an increasing number of the error of secondary task in the appearance implied that subjects felt heavier mental workload. In contrast, the better performance of secondary task meant that the subject felt lower degree of mental workload.

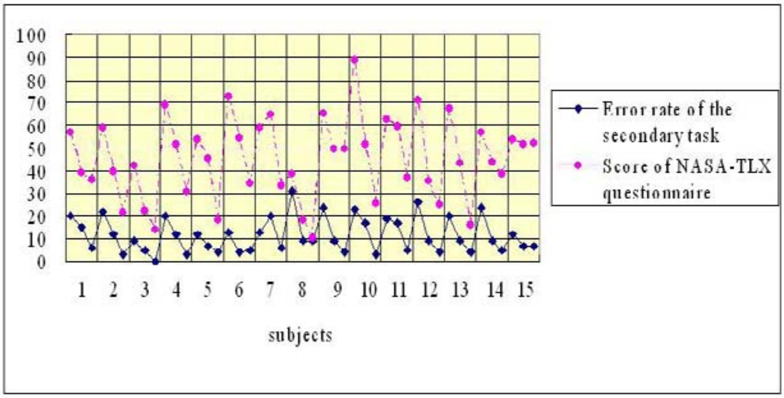


Fig. 3. The relationship between two dependent variables

Table 1. The analysis of Pearson-product moment correlation

Correlations		Error rate of the secondary task	Score of subjective mental workload
Error rate of the secondary task	Pearson Correlation	1	0.691 [a]
	Sig. (2-tailed)	—	0.000
	N	45	45
Score of subjective mental workload	Pearson Correlation	0.691 [a]	1
	Sig. (2-tailed)	0.000	—
	N	45	45

[a] Correlation is significant at the 0.01 level (2-tailed)

3.2 Model Establishment

Seven physiological indexes, including LF/HF (X1), HRV (X2), heart rate (X3), diastolic pressure (X4), systolic pressure (X5), blink frequency (X6), and blink duration (X7) were transferred to the forms of variability ranging from -1 to 1. The accuracy (Y) was to evaluate the performance of the secondary task.

The data X1~X7 were used as inputs in the network system to get an output Y where the output Y is a value from 0 to 100 (percentage of accuracy). Group Method of Data Handling (GMDH) was used to establish model of early warning system.

Thirty-nine data (thirteen subjects) were used to construct the model using NeuroShell software.

The result indicated that all factors of X1~X7 significantly affected the accuracy of the secondary task performance. Also, given the values for X1~X7, a model of performance was yielded and expressed by the following equation:

$$Y = 4.5X_6-33X_4+87-14X_3-0.52X_1+79X_7+590X_3^2-220X_4^2-120X_7^2- 1900X_3^3+1100X_4^3 - 270X_3X_4+840X_3X_7+18X_2^2-34X_6^2- 7.1X_2^3+6.1X_2X_6+130X_4X_7-39X_1X_5 \quad (1)$$

In equation (1), MSE of the model was 9.04 and R square of the model was 0.84. Such equation was expected to provide supervisors/operators a reference value of performance (Y) by giving inputs X1~X7.

For validation, 6 data (2 subjects) were taken into the model. The relational information was described in Table2. In this model, the six estimated values were very close to the real values and all fell in the 95% confidence interval. Therefore, the model was suitable and accurate to estimate the performance.

Table 2. Model validation

Subject Number	X_1	X_2	X_3	X_4	X_5	X_6	X_7	Estimative Value	Real Value	Low Bound of 95% C.I.	Upper Bound of 95% C.I.
14(High)	0.10	-0.61	0.03	0.05	-0.14	0.20	-0.07	83.71	81	77.491	89.929
(Median)	0.00	-0.45	0.02	0.19	-0.12	-0.01	0.02	85.19	83	78.971	91.409
(Low)	0.00	-0.52	-0.05	0.15	-0.17	-0.03	0.00	90.85	95	84.631	97.069
15(High)	-0.33	-0.30	0.00	0.23	-0.04	-0.03	-0.05	77.15	76	70.931	83.369
(Median)	-0.60	-0.22	-0.11	-0.10	-0.05	0.14	-0.03	95.86	91	89.641	102.079
(Low)	-0.67	0.14	-0.15	-0.15	0.01	0.25	0.04	96.44	95	90.221	102.659

4 Discussions

4.1 NASA–TLX Mental Workload and the Performance of Secondary Task

The NASA–TLX subjective mental workload assessment showed a significant correlation with the different level of mental workload of secondary task. For almost all subjects, the highest NASA–TLX scores occurred in the heavy mental workload whereas the lowest scores happened in the low mental workload. This result was not only consisted with the previous studies [15, 16] but also confirmed that the tasks used in this experiment could distinguish the different level of mental workload.

4.2 Physiological Indexes

From literature review, it could be found that some physiological indexes were significantly affected by the mental workload. The physiological indexes measured in this experiment included heart rate, heart rate variability, blood pressure (systolic pressure and diastolic pressure), parasympathetic/sympathetic ratio (LF/HF ratio),

eyes blink frequency and eyes blink duration. The experimental result showed that most of the participants' heart rate and LF/HF components increased when the mental workload increased. On the contrary, the heart rate variety (HRV) decreased when the mental workload increased. These findings were consistent with some previous studies [6; 17]. Aside from these, many participants' blood pressure was not increased as the mental workload increases. This was not consistent with the previous studies [6; 18]. The possible reason could be that the mental workload of this experiment was not high enough to affect the blood pressure change. The experimental result also showed that most of the participants' eyes blink duration was shorter and eyes blink frequency was fewer during high mental workload than during low mental workload. These findings were also similar to many previous studies [17, 18, 19]. The reason could be the participants paid more attention on the interface during high mental workload phase than during low mental workload phase; hence, the activities of eyes (eye blinking frequency and duration) were lower during high mental workload phase.

5 Conclusions

This study developed an early warning model allowing the operators/supervisors to monitor operators' mental workload by physiological indicators. Comparing the synthesized index with the performance level set by supervisor, the early warning system could be commenced to alert the operator while the synthesized index was lower than the setting performance level. For further applications in different fields such as nuclear power plant, flight, aviation, air transportation control, driving and radar vigilance, etc, more data or physiological indicators are needed to train or add to this model.

Acknowledgements. This research has been supported by the National Science Council, Taiwan, Republic of China, Project No. NSC95-NU-7-007-003). The authors wish to thank for the help provided by the Institute of Nuclear Energy Research, Taiwan, Republic of China**.**

References

1. Billings, C.E.: Some questions about advanced automation. In: Mouloua, M., Koonce, J.M. (eds.) Human-automation interaction, pp. 314–320. Lawrence Erlbaum, Mahwah, NJ (1997)
2. Wilson, G.F., Russell, C.A.: Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks. Human Factors 45(4), 635–643 (2003)
3. Moray, N.: Mental workload since 1979. International Review of Ergonomics 2, 123–150 (1988)
4. Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D., Linton, P.M., Wierwille, W.W., Zaklad, A.L., Bittner, A.C.J., Wherry Jr., R.J.: Operator workload: Comprehensive review and evaluation of workload methodologies. US Army Research Institute for the Behavioral and Social Sciences, Willow Grove (1989)

5. Farmer, E., Brownson, A.: Review of workload measurement, Analysis and interpretation methods. European air traffic management programme, CARE-Integra-TRS-130-02-WP2-1-0 (2003)
6. Veltman, J.A., Gaillard, A.W.K.: Physiological indices of workload in a simulated flight task. Biological Psychology 42, 323–342 (1996)
7. Guhe, M., Liao, W., Zhu, Z., Ji, Q., Gray, W.D., Schoelles, M.J.: Non-intrusive measurement of workload in real-time. In: Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society. Orlando, FL (2005)
8. Miyake, S.: Multivariate workload evaluation combining physiological and subjective measures. International Journal of Psychophysiology 40, 233–238 (2001)
9. Chen, D., Vertegaal, R.: Using Mental Load for Managing Interruptions in physiologically attentive user interfaces. In: Proceeding of Conference on Human Factors in Computing Systems, pp. 1513–1516. ACM, New York (2004)
10. Xie, B., Salvendy, G.: Review and reappraisal of modelling and predicting mental workload in single- and multi-task environments. Work. & Stress 14(1), 74–99 (2000)
11. Gevins, A., Smith, M.E.: Detecting transient cognitive impairment with EEG pattern recognition methods. Aviation, Space, and Environmental Medicine 70(10), 1018–1024 (1999)
12. Ivakhnanko, A.G.: Polynomial Theory of Complex Systems. IEEE Transactions on Systems, Man, and Cybernetics SMC 1(4), 363–378 (1971)
13. Pavel, N., Miroslav, S.: Modeling of student's quality by means of GMDH algorithms. Systems Analysis Modeling Simulation 43(10), 1415–1426 (2003)
14. Tomonori, N., Shizue, S.: Identification and forecasting in management systems using the GMDH method. Appl. Math. Modeling 6, 7–15 (1982)
15. Battiste, V., Bortolussi, M.: Transport pilot workload - A comparison of two subjective techniques in Human Factors Society. In: Annual Meeting 32nd. 1989, Human Factors Society, Anaheim, CA, USA (1989)
16. Jones, D.G., Endsley, M.R.: Can real-time probes provide a valid measure of situation awareness? In: Proceedings of the Human Performance, Situation Awareness and Automation (2000)
17. Wilson, G.F.: An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. The. International Journal of Aviation Psychology 12(1), 3–18 (2002(a))
18. Wastell, D.G., Newman, M.: Stress, control and computer system design: a psychophysiological field study. Behaviour & Information on Technology 15(3), 183–192 (1996)
19. Wilson, G.F.: Psychophysiological Test Methods and procedures. In: Charlton, S.G., O'Brien, T.G. (eds.) Handbook of human factors testing and evaluation, pp. 127–156. Lawrence Erlbraum Associates, Mahwah, NJ (2002(b))

Effects of Driver Fatigue Monitoring – An Expert Survey

Katja Karrer and Matthias Roetting

Technische Universität Berlin, Fachgebiet Mensch-Maschine-Systeme, Franklinstr. 28-29, 10587 Berlin, Germany
{Katja.Karrer,Matthias.Roetting}@mms.TU-Berlin.de

Abstract. On long journeys under monotonous road conditions, fatigue monitoring systems might help to reduce sleep-related crashes by informing or alerting the driver, or even by taking corrective actions in the driving task. The objective of this study was to find out more about the view of experts on future driver fatigue monitoring. A questionnaire was designed to discover the objectives and the predicted effects of these systems. Evaluations of 19 researchers and 52 professional drivers were compared to each other. Researchers predict positive effects of fatigue monitoring, as the reduction of accidents, but do not deny possible hazards due to behavioral adaptation. Professional drivers claim it is particularly important to develop an affirmative attitude towards driving without fatigue and see potential in enhancing the individual responsibility of the drivers.

Keywords: driver, fatigue, sleepiness, drowsiness, monitoring, assistance system, expert survey, behavioral adaptation.

1 Introduction

1.1 Monitoring Driver Fatigue

On long journeys under undemanding and monotonous road conditions, sleepiness of the driver poses a serious hazard. Particularly professional drivers with a high driving performance, like long distance lorry drivers, are confronted with the risk of falling asleep at the wheel (e.g. Horne & Reyner, 1999). Driver fatigue (sleepiness, tiredness) is the largest identifiable and preventable cause of road accidents worldwide, accounting for approximately 15-20% of all accidents, with official statistics often underestimating this contribution (consensus statement of an international group of scientists, 2000). The development of new technologies for driver fatigue monitoring is an attempt to reduce sleep-related crashes by informing or alerting the driver, or even by taking corrective actions in the driving task (see European projects like SAVE "System for effective Assessment of the driver state and Vehicle control in Emergency situations" and AWAKE "Assessment of driver vigilance and Warning According to traffic risK Estimation"). These new systems do not only raise crucial questions concerning the technical realization but also in terms of their consequences

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 324–330, 2007.

on driver activity. The introduction of new safety systems does not always result in the desired increase in safety, as has been shown in the past (Hoedemaeker and Brookhuis, 1998; Ward et al., 1995; Rudin-Brown and Noy, 2003; Rudin-Brown and Parker, 2003; Sagberg et al., 1997; Ward et al., 1996). Explanatory models like “risk homeostasis” theory (Wilde, 1994) and “zero-risk” theory (Summala, 1988) imply that any safety system in the car will have a limited effect, due to behavioural adaptation. These adaptations may also appear as a result of the introduction of a fatigue monitoring system and possibly undermine its expected safety benefits. Driver monitorring devices might even encourage sleepy drivers to take further risks and to continue driving.

A lot has been achieved in predicting or detecting driver drowsiness (e.g. SENSATION project). What remains unresolved is 1) how to respond to a drowsy driver and 2) what kind of effect to anticipate with different feedbacks. Regarding the human-machine-interface it is crucial for the success of such a system to find out to what extent interface and feedback have an impact on situational risk perception and behaviour of the driver.

1.2 Objective

The objective of this study was to shed light on “the experts’ view” on driver fatigue monitoring. To find out more about objectives and predicted effects of driver fatigue monitoring, a survey was conducted. In point of fact, the survey had two expert groups. One of the expert groups consisted of researchers working in the field of driver fatigue monitoring. The other group has a different kind of expert knowledge that derives from the direct experience with driver fatigue: professional drivers. We also intended to examine differences between the expert groups.

2 Method

2.1 Questionnaire

A questionnaire was designed to discover the objectives and the predicted effects of driver fatigue monitoring. In addition, demographics like age and gender were collected, as well as measures on expertise.

The main objective of driver fatigue monitoring is to reduce numbers of fatigue-related accidents. Different ways of how this can be accomplished are objectives on a subordinate level. In the first part of the questionnaire these objectives had to be ranked by their relevance by putting a '1' next to the one considered most important, a '2' next to the one next most important, and so on down to '5'. The objectives were: a) activate drivers to manage fatigue while driving, b) adapt the thresholds of other assistance- and safety-systems (such as adaptive cruise control) depending on the fatigue level, c) develop an affirmative attitude towards driving without fatigue, d) educate drivers about the signs of fatigue, e) improve drivers’ awareness of the risks of fatigue while driving, f) increase drivers’ general understanding of fatigue

development, g) interfere actively in the driving performance to prevent accidents, h) make drivers aware of their own current fatigue level, i) motivate fatigued drivers to take a break, j) warn drivers, before their driving performance decreases seriously or k) another important objective, which could be specified.

In the second part of the questionnaire possible positive and negative outcomes of driver fatigue monitoring were listed as statements. The outcomes were:

- Drivers will have better fatigue awareness with such a device.
- Drivers will cause less road accidents with such a device.
- Drivers will drive more responsibly with such a device.
- Drivers will improve the ability of self-monitoring with such a device.
- Drivers will take more breaks when using such a device.
- Drivers will estimate the risk of driving fatigued more adequately.
- Drivers will drive more safely with such a device.
- Drivers will underestimate the risk of driving fatigued when using such a device.
- Drivers will care less about self-monitoring fatigue with such a device.
- Drivers will overestimate their own driving ability when using such a device.
- Drivers will tend to overtrust such a device.
- Drivers will tend to leave fatigue control to such a device.
- Drivers will be additionally strained by such a device.
- Drivers will be distracted by such a device.
- Drivers will drive longer with such a device.
- Drivers will feel safer with such a device.

The extent to which the respondent agreed or disagreed with the statements had to be indicated on a 5-point rating scale (-2 “strongly disagree”, -1 “disagree”, 0 “neutral”, 1 “agree”, 2 “strongly agree”). The statements related to the effects of three different feedback types which differed in their extent of automation. Respondents were asked to imagine a driver-fatigue monitoring device which 1) interferes actively in the driving performance when the driver falls asleep, 2) warns the driver, before his or her driving performance decreases seriously and 3) continuously informs the driver about his fatigue level. The survey took about 20 minutes to complete.

2.2 Participants

49 questionnaires were sent to researchers all over the world working in the field of driver fatigue monitoring. 23 returned questionnaires were checked for expert status using the criterion of working more than 3 years on the topic. 19 returned forms could be considered. In the group of professional drivers, a total of 52 replies from online and paper forms were analysed.

2.3 Statistical Analyses

All questionnaires were entered into the statistics software SPSS 14.0. Estimations and prognoses of both groups were compared to each other.

To examine the rank order of objectives contributing to the main goal of reducing fatigue-related road accidents, points were assigned according to their rated relevance. The objective ranked as the fifth most important factor received one point, and the objective considered as most relevant received five points, the objectives in between were handled accordingly. Zero points were assigned to non-selected objectives. Average points per objective determined the final rank order.

Rating scales concerning outcomes of driver fatigue monitoring were analysed using an analysis of variance with the within-subjects variable "type of feedback" (interfere / warn / inform) and the between factor "expert group" (research experts / professional drivers). The 0.05 level of significance was adopted for these analyses.

3 Results

3.1 Characteristics of Experts Who Completed Questionnaires

Of the 19 research experts who returned completed questionnaires, 17 were men. Their mean age was 50 years (range 30–81 years). Six identified themselves as citizen of USA, three from Sweden, Netherlands and Germany, and one from France, Canada and Australia. The amount of years that experts reported working as researchers in the field of driver monitoring was more than 10 years for ten participants, and 3 to 10 years for the nine remaining.

49 of the 52 professional drivers were men; except for 3 of them all were German citizens. The mean age was 42 years (range 24–62 years). Approximately 63,5% of the professional drivers were working in truckage companies, 19,2% in bus companies, 11,5% in another sector, 5,8% were not specified. They drove annually an average of 120.000 kilometers (range 20.000-250.000).

3.2 Evaluation of Objectives

Table 1 presents the results for both expert groups. The most important objective from the researchers' view is to warn drivers before their driving performance decreases seriously (2.95), and to make drivers aware of their own current fatigue level (2.37). In comparison to professional drivers, they attach more importance to objectives like activating the driver (1.11) or adapting thresholds of other systems (1.21). Professional drivers see the most promising objective in motivating drivers to take a break (2.15) and improving drivers' awareness of the risks of fatigue while driving (1.94). They also differ from researchers in ascribing more importance to increasing the general understanding of fatigue development (1.08) or to developing an affirmative attitude towards driving without fatigue (1.60). Both groups agree in attaching importance to motivating drivers to take a break (2.11; 2.15).

Table 1. Mean score for objectives according to researchers and professional drivers

Objectives	Researchers	Professional Drivers
Warn drivers, before their driving performance decreases seriously	2.95	1.83
Make drivers aware of their own current fatigue level	2.37	1.67
Activate drivers to manage fatigue while driving	1.11	0.69
Adapt the thresholds of other assistance- and safety-systems (such as ACC) depending on the fatigue level	1.21	1.04
Interfere actively in the driving performance to prevent accidents	1.79	1.69
Motivate drivers to take a break	2.11	2.15
Develop an affirmative attitude towards driving without fatigue	1.26	1.60
Increase drivers' general understanding of fatigue development	0.47	1.08
Educate drivers about the signs of fatigue	1.05	1.69
Improve drivers' awareness of the risks of fatigue while driving	1.16	1.94

3.3 Feedback Effects

The ANOVA indicated some significant differences between the predicted effects of the three feedback types (interfere / warn / inform). There are concerns in both groups that drivers will underestimate the risk of driving fatigued primarily when using an interfering device (see table 2). For researchers, the danger of driving longer than before is related to the degree of automation of the system, as it is considered highest with the interfering device. Professional drivers judge the interfering system as most promising to reduce accident numbers. Researchers favour the warning feedback over the interfering system. In comparison, a continuously informing device is seen as most suitable to enhance fatigue awareness and self-monitoring. Yet, strain and distraction is also seen as highest with this kind of feedback. All reported differences were significant on the 0.05 level.

There were also significant differences between the judgements of both expert groups. Research experts had generally a much more optimistic view on driver fatigue monitoring systems than professional drivers. They assumed better fatigue awareness, fewer accidents, more adequate risk estimation, improved self-monitoring and more breaks when implementing such systems than the drivers. However, they also saw more risk of driving longer than professional drivers. All these reported differences were significant on the 0.05 level.

Table 2. Mean ratings for expert group and feedback type (scale from -2 to 2)

	Expert group	Researcher			Professional Driver		
Drivers will …	Feedback type	Interfere	Warn	Inform	Interfere	Warn	Inform
… underestimate the risk of driving fatigued when using such a device.		0.47	0.11	-0.35	0.49	-0.08	-0.20
… have better fatigue awareness with such a device.		0.63	0.89	1.24	-0.39	-0.31	-0.13
… be additionally strained by such a device.		-0.74	-0.50	0.06	-0.47	-0.67	0.04
… be distracted by such a device.		-0.53	-0.17	0.25	-0.59	-0.57	0.07
… care less about self-monitoring fatigue with such a device.		0.47	0.61	0.24	0.16	0.17	0.19
… cause less road accidents with such a device.		0.90	1.00	0.65	0.15	-0.09	-0.33
… drive longer with such a device.		0.95	0.67	0.12	-0.59	-0.40	-0.43
… drive more responsibly with such a device.		-0.32	-0.17	0.12	-0.39	-0.43	-0.34
… drive more safely with such a device.		0.42	0.61	0.24	-0.18	-0.22	-0.24
… estimate the risk of driving fatigued more adequately.		0.05	0.11	0.77	-0.37	-0.39	-0.45
… feel safer with such a device.		1.00	1.00	0.82	0.24	0.39	0.29
… improve the ability of self-monitoring with such a device.		0.26	0.72	0.82	-0.50	-0.42	-0.34
… overestimate their own driving ability when using such a device.		0.37	0.39	0.00	-0.32	-0.33	-0.33
… tend to leave fatigue control to such a device.		0.37	0.44	0.29	-0.02	-0.04	0.04
… tend to overtrust such a device.		0.58	0.56	0.47	-0.04	-0.06	-0.04
… take more breaks when using such a device.		0.05	0.06	0.12	-0.66	-0.63	-0.75

4 Discussion

The results of this study demonstrate that researchers working in the field of driver fatigue monitoring see potential in a system which informs and warns drivers, but also activates them and adapts thresholds of other systems. The preferred feedback is more a correction of critical states and actions. Professional drivers place more emphasis on the individual responsibility of the drivers, on motivating them to take a break, improving their awareness of the risks of fatigue while driving, and even on changing drivers' attitude towards driving without fatigue. Here the preferred approach deals more with changing misinterpretations and a reeducation of drivers. The results of the objective ranking might help improve the implementation strategies of such devices. From the view of user-centred design it is important for the developers and designers to give attention to the opinion, the prospects, and the objections of the end user of the

systems' feedback. Experienced drivers seem to value educative campaigns somewhat higher than a direct intervention in their driving task.

Both expert groups agree that an interfering system might lead to an underestimation of the risk of driving fatigued. Indeed, the researchers are quite aware of possible behavioural adaptations, like leaving fatigue control to the system, overestimating the own driving ability or overtrusting the system.

But overall, researchers believe that drivers will cause less road accidents having a driver monitoring device on-board. The general optimism of researchers concerning positive effects of such systems is not shared to the same extent by professional drivers. The latter do not believe that systems might help improve the ability of self-monitoring or even cause drivers to take more breaks. Possibly, they do not see the problem in incorrect fatigue self-assessment but rather in an insufficient responsibility of the drivers. Analysing and foreseeing how drivers are likely to use the system might help to avoid effects such as inducing sleepy drivers to prolong the driving and therefore to fail at reducing accident risk. Future research should investigate how the feedback could be designed to appeal to the drivers' individual responsibility rather than correcting their self-assessment and actions.

References

1. Consensus Statement: Fatigue and accidents in transport operations. J. Sleep Res. 9, 395 (2000)
2. Hoedemaeker, M., Brookhuis, K.: Behavioural adaptation to driving with an adaptive cruise control (ACC). Transportation Research F 1, 95–106 (1998)
3. Horne, J., Reyner, L.: Vehicle accidents related to sleep: a review. Occupational and Environmental Medicine 56, 289–294 (1999)
4. Rudin-Brown, C.M., Noy, I.: Investigation of Behavioral Adaptation to Lane Departure Warning. Transportation Research Record 1803, 30–37 (2003)
5. Rudin-Brown, C., Parker, H.: Behavioural adaptation to adaptive cruise control (ACC): implications for preventive strategies. Transportation Research F 7, 59–76 (2003)
6. Sagberg, F., Fosser, S., Saetermo, I.A.: An Investigation of Behavioural Adaptation to Airbags and Antilock Brakes Among Taxi Drivers. Accid. Anal. And Prev. 29(3), 293–302 (1997)
7. Summala, H.: Risk control is not risk adjustment: the zero-risk theory of driver behavior and its implications. Ergonomics 31(4), 491–504 (1988)
8. Ward, N. J., Humphreys, M., Fairclough S.: A field study of behavioural adaptation with an autonomous intelligent cruise control system. In: Handbook of the International Conference on Traffic and Transport Psychology, 22-25 May, Valencia, Spain, pp. 15–19 (1996)
9. Wilde, G.J.S.: Target risk. PDE Publications, Toronto Canada (1994)

Study on the Instruction Method for Plant Operator

Daiji Kobayashi[1], Hiroaki Murata[2], and Sakae Yamamoto[2]

[1] Nihonbashi Gakkan University
1225-6 Kashiwa Kashiwa-city Chiba,
277-0005 Japan
[2] Tokyo University of Science,
1-3 Kagurazaka Shinjuku-ku Tokyo,
162-8601 Japan

Abstract. In this study, the characteristics of three training curriculums were compared from the viewpoint of the trainee's cognitive process. In the experiment, the nine participants set into three groups of three participants, and trained in each curriculum. In order to evaluate the trainee's cognitive behavior in identifying malfunctions, the concept of mental algorithm was used. As the results, the trainee's cognitive process for identifying malfunctions is estimated as a model. This model could reflect the contents of training curriculum regarding the trainee's cognitive process.

Keywords: plant, operator, instruction, thinking process.

1 Background

For a plant operator, it is important to find out the cause of plant disturbance. When the plant's components have some trouble, the plant operators have to identify the troubled situation according to some procedural manuals, documented instructions and so on. However, our experimental studies indicated that the operator's empirical techniques affect their subtle thinking for identifying malfunctions. Kobayashi (2003) investigated the thinking process including subtle thinking through some simulator experiment, and named the characteristic thinking process *mental algorithm*. The *mental algorithm* is represented by the sequence of subtle thinking steps. Each step of subtle thinking, occurs between getting information from the interface, is mainly composed of *cognitive state* and *strategy*. Therefore, the transition of the cognitive state and the contents of strategy could represent the operator's cognitive behavior. From this point, Kobayashi evaluated some types of plant interfaces (the screens on CRT display) experimentally.

The plant operator is used to learn the techniques for the practical operation task in accordance with their experiences. The techniques, of course, are not only for appropriate action but also for cognitive behavior. The cognitive behavior which would be observed by the mental algorithm is the product of experiences, instructions, trainings, and so on. In case of Japanese nuclear power plant, the training curriculums for the operators are systematized through trial and error to some extent. The reason

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 331–338, 2007.

why it is difficult to pursue more appropriate training curriculum is based on two facts owing to the variation of the plant and the human operator as follows:

First, some types of nuclear power plant has been developed with the advancement of technology in Japan, then some types of full-scale simulator has been used for operator training. In case of BWR power plant, there are three types of BWR power plants, and three types of full-scale simulator are used. Moreover, the specifications of the same type of the BWR power plants are slightly different each other, therefore the trainer have to take into account the variation especially for the simulator training.

Next, it is worried about to train up the successors for old plants, because Japanese face the serious concerns of rapid aging and very low birthrate. Many veteran operators, are baby boomers, reach the retirement age, have great deal of experiences and skills. These experiences and skills could not represent by words, therefore it is difficult to hand their experiences and skills on the next generation as early as possible.

Concerning about the above facts, the operator training curriculums for BWR power plant should be re-considered based on their appropriateness.

Previous studies proposed some model representing the relation between the operator's skill and their cognitive behavior. For example, Rasmussen's SRK-model is a pioneering work, and GEMS dynamic model proposed by Embrey and Reason (1986) is concretized and detailed SRK-model. On the other hand, Yukimachi and Hasegawa (1999) revised SRK-model in order to analyze the human errors at nuclear power plant in Japan. In these models, the relation between the operator's skill level and cognitive behavior is represented based on some observational studies and example analysis of human errors. However, it is difficult to use that knowledge to make the training curriculum, and we tried another approach to the evaluation of the training curriculum from cognitive viewpoint.

Although our goal is to develop the method of evaluation for instruction and training curriculum, we tried to investigate the relation between training curriculum and the characteristics of the operator's thinking process for identifying malfunctions as the first step of our research. Therefore, the object of this paper is to reveal the cognitive mechanism of the instruction, and the training effect on operator's thinking process for identifying malfunctions.

2 Method

In order to compare the cognitive processes of each participant, we conducted the same simulator experiment after training by different three curriculums of each groups. In the experiment, the number of cognitive failure and the types of the cognitive failure were observed based on their performances and verbal protocols.

2.1 Apparatus

The participants' verbal protocols and their performances were recorded by a video camera, and the participant's eye tracking was recorded by an eye mark analyzer (NAC Image Tech. EMR-8) using a simulator after the training. The simulator was composed of PC and 22-inches touch monitor (see Figure 1). Figure 2 shows the

screen of the touch monitor indicating the part of a nuclear power plant's process, 12 annunciators, and 2 indicators for selected component. The indicators on the lower part of the screen are a line chart and a parameter list. These are updated every second. In addition, each component's main parameter on the upper part is also updated every second. When a participant wants to refer a component's all parameters, the participant has to select and touch the component icon on the screen using a stick. The simulator records these historical operation data in its hard disk.

2.2 Participants

Participants were nine male Japanese university students. They did not have any knowledge about the plant operation; however they have the knowledge about basic thermodynamics. In order to compare the participant's cognitive behavior in simulator experiment, we set the nine participants into three group (Group-A, -B, -C) of three participants, and trained in each curriculum.

2.3 Training

Firstly, an experimenter lectured all participants on the way of simulator operation and basic knowledge about the process for supervising. Mainly we told a behavior and dynamics of components.

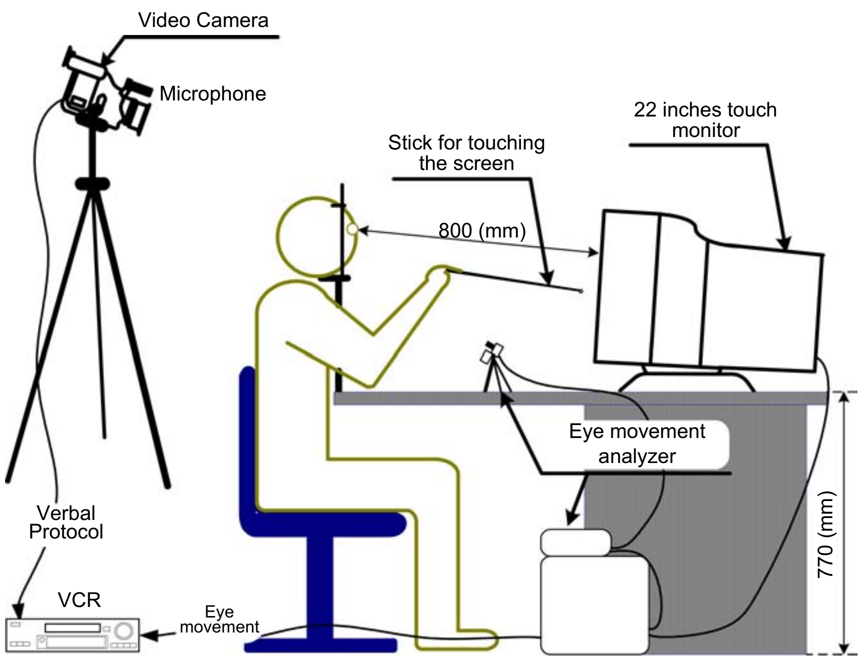


Fig. 1. Experimental apparatus

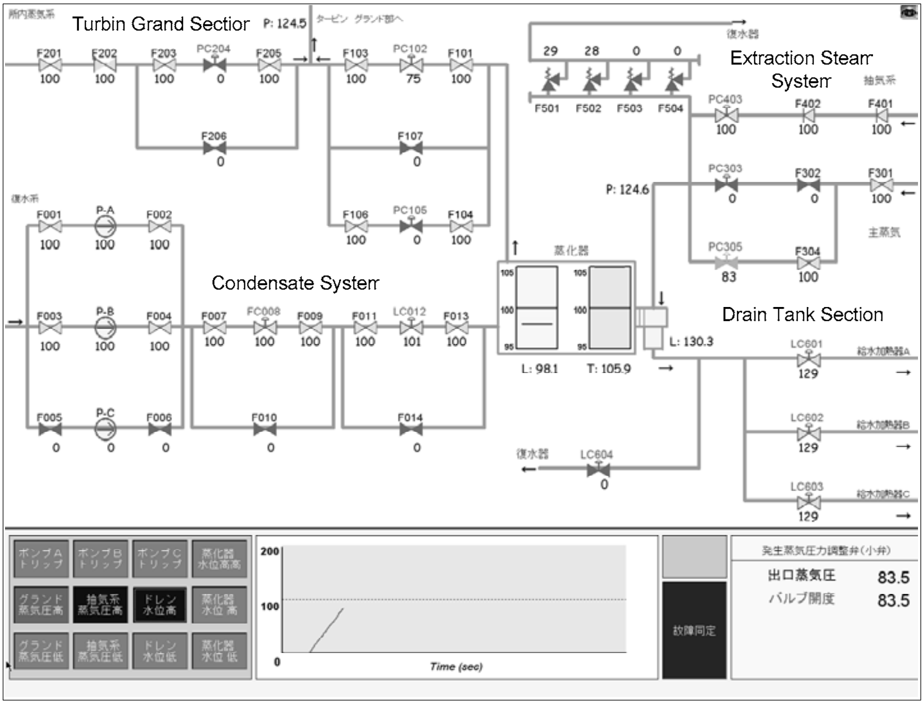


Fig. 2. Contents of the simulator’s screen on CRT

Next, the experimenter instructed how a malfunction affects the other components’ parameter.

Last, the experimenter trained the participants in the following different way: In case of Group-A, the participants underwent simulator training using ten scenarios including a malfunction of each; To the participants belonging to Group-B, the experimenter lectured on the empirical way of identifying malfunctions in order to make them obtain some rule-based way of thinking. After the lecture, they underwent same simulator training; In case of Group-C, we instructed the participants to think or image the transitional parameter of some components after a malfunction occurred in order to make them have the image of each component's transitional state. After the instruction, they underwent the simulator training in the same way.

2.4 Simulator Experiment

After the simulator trainings, we conducted simulator experiments using fifteen scenarios including a few malfunctions. Each participant tried to identify all malfunctions in each scenario, and tried to speak their thinking process while a trial. Their performance and verbal protocols were recorded by a video camera, and the component’s name and the time they referred was recorded by PC. After the experiment, we heard about the participant’s opinion about the trials.

Table 1. Comparison of curriculum contents for three groups

Curriculum contents	Group		
	A	B	C
Instruction by lecturing using slides on the followings: ・Outline of the process ・Each component behavior in the processes	1st Day	1st Day	1st Day
Paper test for the understanding of the above lecture	2nd Day	2nd Day	2nd Day
Instruction on the diagnosis methods from the component behavior using the simulator and a manual	N/A	3rd Day	N/A
Demonstration of the identifying malfunctions using the simulator	3rd Day		3rd Day
Training of the reasoning skill in estimation of malfunctions using the simulator' s still screen	N/A	N/A	
Training of identifying malfunctions included 10 scenarios using the simulator	4th Day	4th Day	4th Day

N/A: not applicable

3 Result

As the consequence of analyzing the participants' verbal protocols, some cognitive failures in their thinking process were found. The cognitive failures were, for example, misunderstanding, decision error, overlooking, and so on.

3.1 Cognitive Failure

The averaged number of cognitive failure was 8.3 per participant of the Group A. The most of these cognitive failures were to be confused in determination of the component's state. Therefore, the failures could be caused by the lack of thinking skill for identifying malfunction.

The Group B's averaged number of cognitive failures was 4.7 per participant. Most of the cognitive failures were occurred by "one-track way thinking" which would be caused by using inappropriate rules for thinking.

The Group C's averaged number of cognitive errors was 3.0.per participant. Most of the cognitive failures were occurred by careless oversight or perceived notion. In addition, it was found that a participant was not able to acquire the way to identify malfunction sufficiently.

The contents of participants' cognitive failure could be categorized as shown in Table 2.

Table 2 indicates that the cognitive failure by the participants of Group A occurred more frequently than the other groups' cognitive failures, especially the participants of Group-A tended to fail on the stage of grasping the situation in the identification process.

Table 2. Cognitive failures of each participant in each group

Cognitive failure \ Group / Perticipant	A: a	A: b	A: c	B: d	B: e	B: f	C: g	C: h	C: I
Identification process: Oversights of parameter changes indicating the unusual process condition	4	3	1	2	2	0	1	1	0
Identification process: Mistaking component' s conditions	2	2	3	2	0	2	1	1	0
Identification process: Mistaken decision about malfunction	2	1	1	1	2	2	2	1	2
Identification process: Confusion on deducting the state of the process from the components' condition	2	4	0	0	0	1	0	0	0
Total	10	10	5	5	4	5	3	3	2

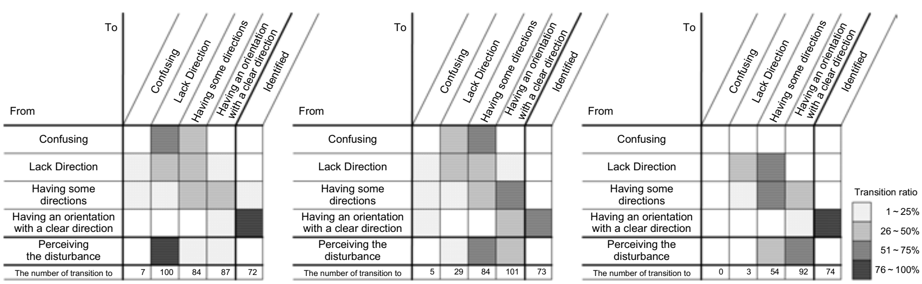
3.2 Mental Algorithm

In order to investigate the reason why the frequency of cognitive failure varies, the mental algorithm of participants in the simulator experiment was estimated based on their verbal protocols and performances. Previous studies evaluated the operator's thinking process based on the transition of cognitive states from their perception about the plant disturbance to their identification about the malfunctions of those. Therefore, the reason of the cognitive failures was investigated through the transition of cognitive state.

From the result of the simulator experiment, we found that the transition of the cognitive state occurred whenever the operator gets the component's state by touching the component's icon on screen. From this point, we estimated the cognitive state, counted the transition pattern of the cognitive state on a group, and calculated the transition pattern's ratio of each group. As the result, four types, that are *confusing*, *lack direction*, *having some directions*, and *having an orientation with a clear direction*, categorized as the cognitive states.

Table 3 shows the transition pattern's ratio of the cognitive state and the number of transition to each cognitive state. The number of transition is the number of getting a component state. Therefore, Table 3 indicates the Group-A's participants search more

Table 3. Transition pattern of cognitive state in each group (left side: Group-A, middle: Group-B, right side: Group-C)



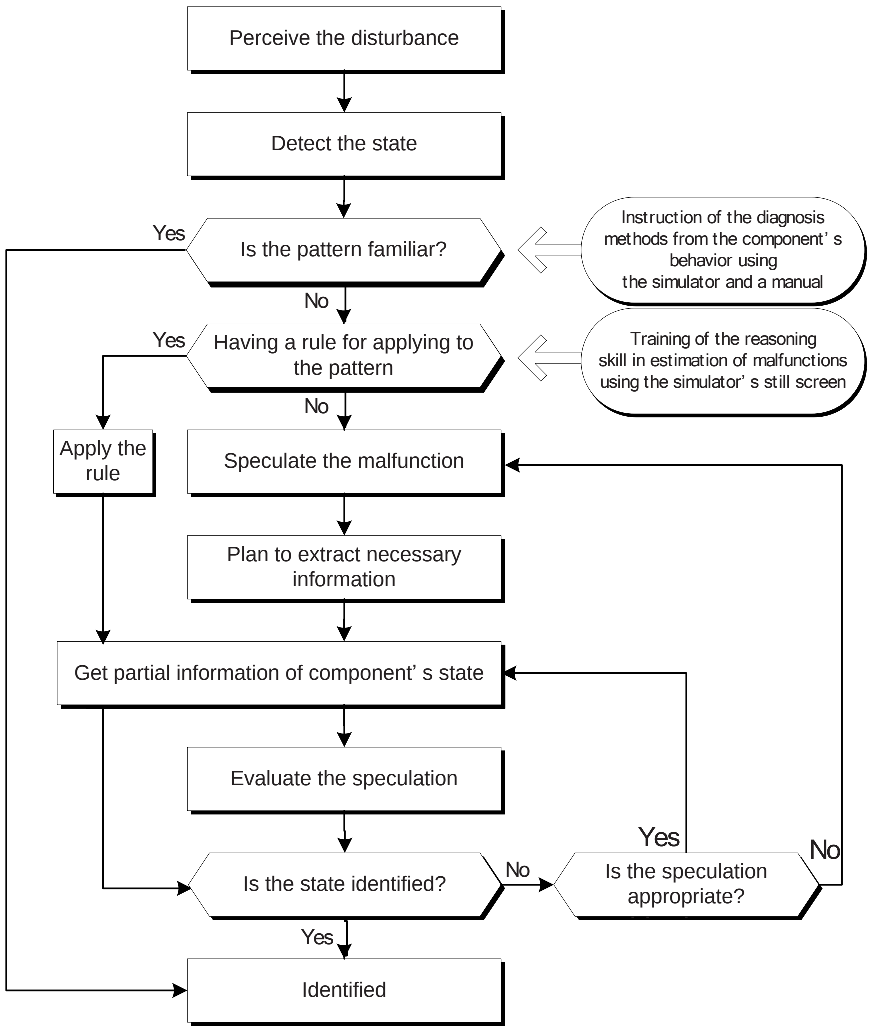


Fig. 3. Trainee's cognitive model for identifying malfunctions

components than the others. Further, the number of transition to *confusing* or *lack direction* is also more than the others are. This means the Group-A's participants could not obtain information for identifying malfunctions in spite of the trainings.

In case of Group-B, the number of transition is higher than Group-C's, and transition to *confusing* and *lack direction* is more frequently than Group-C's. Therefore, their identifying process could not be sufficient.

From these results, we found that the Group-C's cognitive condition is the best cognitive condition in three groups we tried.

4 Discussion

The results of the mental algorithm analysis indicates that Group-A's participants could not obtain the way of identifying malfunctions sufficiently. Therefore, the number of cognitive failures would be more than the other groups'. In other words, it is necessary for the Group-A to learn the diagnosis methods from the component's behavior on screen.

The group-C's participants could not speculate logically but could depend on our rules that we had instructed in advance. Therefore, these results represent that it is important for the operators to understand the mechanism of plant's process.

From above of these investigations, the trainee's cognitive process for identifying malfunctions is estimated as Figure 3. This model reflects the contents of training curriculum regarding the trainee's cognitive process.

5 Conclusion

We compared three types of the training methods and investigated it from cognitive process. Consequently, we could indicate the relation between the characteristics of these trainings and the cognitive process.

References

1. Kobayashi, D., Yamamoto, S.: Experimental study of evaluating graphical CRT displays in process plants based on operator's thinking process. The Journal of Japan Ergonomics Society, Vol. 37, No. 3, 113-124 (2001)
2. Sakata, K., Nishitani, H.: Considerations on Simulator Training for Plant Operations during Malfunctions. The Journal of the Society for Industrial Plant Human Factors of Japan, Vol. 5, No. 2, 130-140 (2000)
3. Sheridan, T.B.: HCI in Supervisory Control: Twelve Dilemmas. Lecture Notes in Control and Information Sciences, Vol. 253, 1-12 (2000)
4. Stammers, R.B.: Training issues: Computer Science Today. Human Factors in Nuclear Safety, 189-196 (1996)
5. Yukimachi, T., Hasegawa, T.: Analysis of human errors at nuclear power plants in Japan. The Journal of the Society for Industrial Plant Human Factors of Japan, Vol. 4, No. 1, 48-57 (1999)

Examining the Moderating Effect of Workload on Controller Task Distribution

Paul U. Lee, Joey Mercer, and Todd Callantine

San Jose State University Foundation / NASA Ames Research Center, Mail Stop 262-4, Moffett Field, CA 94035
{plee,jmercer,tcallantine}@mail.arc.nasa.gov

Abstract. Efforts to characterize controller workload – a key factor in limiting en route capacity – have produced mixed results. Subjective workload ratings reveal significant variations in minimum/maximum workload across individuals and show a categorical jump in perceived workload with a linear increase in aircraft count, making it difficult to predict workload limits with increased traffic. In addition, workload seems to be actively moderated by the controller to reduce monitoring tasks during high traffic/workload situations. In this paper, we examine this strategy shift by associating bookkeeping tasks and route/altitude clearances with online workload ratings. Overall, the data suggest that the controllers shed peripheral tasks related to monitoring and bookkeeping as the traffic ramps up and their perceived workload transitions from low to high. Whenever workload reached a maximum, some bookkeeping tasks were delayed and performed in "groups" after the peak traffic subsided.

Keywords: workload, task load, air traffic control, non-linear, controller strategy, situation awareness.

1 Introduction

In developing future Next Generation Air Transportation Systems (NextGen), controller workload has been identified as a key limiting factor for a significant increase in capacity and therefore has been an active field of research (e.g. [1],[2],[3]). Because workload ratings are subjective and highly prone to individual differences, some researchers have tried to replace workload with more objective metrics, such as aircraft count, number of altitude changes, number of coordination events, traffic geometry, total time in sector, etc. In related research the collective effect of all factors that contribute to air traffic control complexity was examined and termed "dynamic density" [4]. One of the key motivations for dynamic density research is to find a set of metrics that can replace current day Monitor Alert Parameters (MAPs) to predict traffic complexity and associated controller workload.

A significant caveat in substituting these objective task load metrics for workload ratings is that their relationships are often non-linear. For example, we conducted a "traffic load test" in which this non-linear relationship between workload and aircraft count was examined by asking controllers to provide feedback on the maximum

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 339–348, 2007.

traffic they can handle and then adding a few more aircraft to see their reactions [5]. We asked controller participants to manage the challenging traffic scenarios and then increased or decreased the peak aircraft count depending on their workload feedback. The consistent results from the three sectors across participants were that adding only a few aircraft above the peak caused perceived workload to escalate from moderate to high and then to unmanageable. The left graph in Fig. 1 illustrates this point for one of the test sectors. The ten-minute peak aircraft count was 17.2, 19.9, and 22.7 aircraft for moderate, maximum, and unmanageable workload, respectively, suggesting that minor changes in peak aircraft count had significant variations in perceived workload.

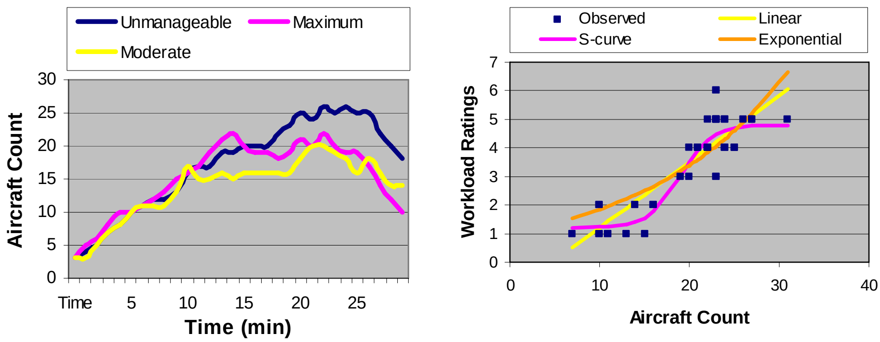


Fig. 1. Aircraft count in different traffic scenarios with moderate, maximum, and unmanageable workload during the traffic load test *(left)*; Scatter plots of workload vs. aircraft count with linear, exponential, and S-curve regression fits *(right)*

Since the peak aircraft count has emerged as one of the best predictors of workload [6], we examined the relationship between workload and aircraft count in more detail using another set of data. During this study controller participants reported their workload every five minutes during the simulation runs on a scale of 1 to 7 using a Workload Assessment Keyboard (WAK) [3]. Workload ratings were then correlated with peak aircraft count at each five minute interval. As shown in the right graph in Fig. 1, reported workload was low for aircraft count up to 16 and then quickly ramped up to high workload from 16 to 22 aircraft. Linear, exponential, and S-curve (estimating a categorical jump) functions were modeled and tested to see which function provided the best fit to the observed data. The results showed that the S-curve function captured more of the variance more consistently across multiple sectors and participants, suggesting that controllers perceive workload increases categorically with respect to the traffic count [7]. This categorical perception of workload matches well with a general phenomenon that controllers often report a low to moderate level of workload for seemingly busy traffic but report much higher workload with a few added tasks and/or minor off-nominal events once a certain traffic level is reached. In general, the ratings of different controllers across different test scenarios resulted in significant variations in minimum/maximum workload ratings as well as the shape of the curve. Given such a large variability, it would not be prudent to consider workload ratings as raw indicators of the maximum traffic that an "average" controller can handle since that number would vary significantly per

controller. However, a key advantage of the S-curve functions was that they modeled the data more consistently across different individuals [7], suggesting that S-curve functions best capture the underlying cognitive perception of perceived workload.

2 Effect of Workload on the Controller Tasks

Another challenge in correlating workload directly with task load metrics is that workload may be a causal factor that drives the task load distribution. In other words, workload is often an input factor that controllers use to actively moderate their task load distribution in order to cap one's own workload at an acceptable level. For example, controllers may try to provide efficient routings during low traffic scenarios, even if it requires more workload but abandon such practices during high traffic/workload situations in favor of actions that minimize workload.

To examine the moderating effect of workload on the types and the number of controller tasks, we re-analyzed the data from the two studies mentioned above. First, the task load – i.e. frequencies and types of controller tasks – was examined and compared with the aircraft count from the traffic load test, then the task load was compared with the associated workload ratings from the second study in a time-series plot. The results from these analyses are described below.

2.1 Task Distribution in the Traffic Load Test

Task load metrics were divided into three main categories: handoffs, clearances, and monitoring tasks. Task load was analyzed for the three test sectors, but results here will focus on a particular sector due to space limitations. Overall, the pattern of results was similar across the three sectors.

The number of handoffs that a controller accepts from an upstream sector and initiates to a downstream sector is directly related to number of aircraft in their sector. The handoff-related tasks therefore mirrored the aircraft count. Controllers also engaged in various monitoring tasks. Most of the monitoring tasks were not recorded by the data collection system, but the ones that were show some interesting patterns. The left graph in Fig. 2 shows the average number of times per aircraft that the controller participants toggled or adjusted the datablocks, displayed FMS routes, and displayed J-rings around the targets. Datablock adjustments were often used as memory aids to let the controllers visually discriminate between aircraft that have been handed off, need to be attended to, etc. Toggles were used to minimize a datablock once aircraft were handed off to the next sector. Displaying FMS routes allowed the controllers to verify where the planes were going, which was important because the airspace and the traffic scenarios were unfamiliar to them. J-rings were often used to visually emphasize the 5 nm separation boundaries for aircraft that had potential conflicts with other nearby aircraft.

An interesting pattern in this graph is that as the traffic increased, these types of monitoring tasks decreased, suggesting that as controllers became busier with traffic management duties, the number of lower priority monitoring and/or bookkeeping tasks decreased.

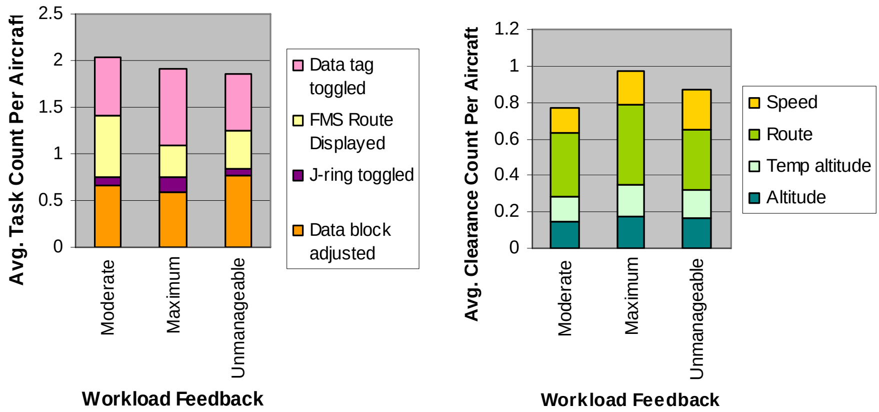


Fig. 2. Number of monitoring-related tasks (*left*) and number of clearances (*right*) per aircraft

In contrast, the number of clearances demonstrated a different pattern. The right graph in Fig. 2 shows the number of speed and route clearances controllers issued using data link, as well as the number of altitude clearances issued by voice. These results show that the number of clearances increased as the traffic increased from “moderate” to “maximum”, suggesting that a slight increase in traffic count in this case created a significant increase in either the actual or perceived complexity of the traffic, leading the controllers to issue more clearances per aircraft. Interestingly enough, in the “unmanageable” traffic level, the overall number of clearances per aircraft was reduced from the “maximum” traffic level. General observations and controller feedback from the simulation runs suggest that the reduction is perhaps due to controllers abandoning clearances related to efficient traffic management to focus solely on separation related clearances.

3 Time Series Analyses of Workload vs. Task Load

Different patterns emerge with increasing traffic for the three types of task load identified in the study. Handoff-related tasks and the associated workload generally increase with aircraft count. In contrast, as the workload reaches its maximum, controllers seem to shed other bookkeeping and monitoring tasks while still trying to maintain efficient traffic flow management. However, when the traffic becomes unmanageable, they abandon efficient flow management as well and simply try to “survive” by maintaining adequate separations. By effectively shedding lower priority tasks as the traffic increases, the controllers seem to effectively keep their workload below the maximum threshold. To further examine this hypothesis, we examined the other data set to see if there is a shift in the task distribution as the subjective workload increases.

3.1 Method

For these analyses, we used high traffic simulation data that were collected during an evaluation of the En Route Free Maneuvering concept element in NASA’s Distributed

Air-Ground Traffic Management (DAG-TM) project [8]. DAG-TM studies were conducted with an assumption of far-term equipage levels, including fully integrated advanced air and ground decision support tools (DSTs) with data link. The aircraft in this environment were assumed to always fly with 4-D trajectory intent which was probed for conflicts along the 4-D paths. In addition, the ground-side DSTs had 4-D route and altitude trial planning capability, which allowed the controllers to graphically construct a new conflict-free path that could be sent to the flight deck via data link. Transfer of communication was also automated and integrated with data link, sending the frequency change uplink message to the flight deck with the handoff acceptance of the next sector.

There were four experimental conditions in this study, two of which will be re-analyzed and discussed in this paper. In the first "high traffic" condition, the aircraft count reached a peak that was higher than current day MAP values, resulting in a peak workload level that was beyond the allowable level in current day operations. In a second "moderate traffic" condition, the peak aircraft count was reduced by approximately five aircraft, resulting in a comfortable moderate workload for the controller participants. Four data collection runs were analyzed per condition.

Subjective workload assessments were collected from controllers with the WAK using the Air Traffic Workload Input Technique (ATWIT) [3]. Controllers were required to rate their workload on a scale of 1 to 7 at 5-minute intervals throughout each simulation run. In the same five minute intervals, various task load metrics, such as clearances, and handoffs, as well as the average aircraft count during the interval were tabulated. The results are described in the following section.

3.2 Results

Workload vs. Aircraft Count. Fig. 3 illustrates the traffic pattern for an en route sector during the high and moderate traffic conditions, which shows that the peak aircraft count (averaged across 5-minute time span) reached 23 and 17 aircraft respectively.

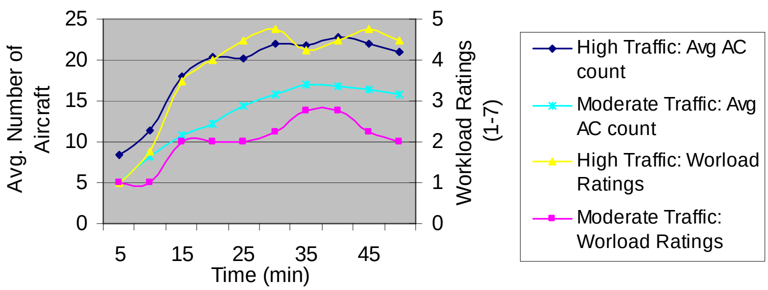


Fig. 3. Number of aircraft count vs. workload ratings for high and moderate traffic scenarios

Similar to the traffic load test results, a difference of five aircraft between the two traffic scenarios resulted in large differences in workload ratings. High traffic scenarios resulted in average workload ratings between 4 and 5, which corresponds to high workload for most controllers, as they generally reserve 6 and 7 ratings to report

situations with catastrophic failures or major re-planning such as in heavy thunderstorm or airport closures. In contrast, moderate traffic scenarios resulted in workload ratings between 2 and 3, which correspond to low to moderate workload.

Workload vs. Monitoring/Bookkeeping Tasks. Similar to task load analyses for the traffic load test, the tasks were divided into handoff-related, clearances, and monitoring/bookkeeping tasks. In this section, we focus on the monitoring/ bookkeeping tasks because our main hypothesis was that the controllers would shed less essential tasks as their workload increased.

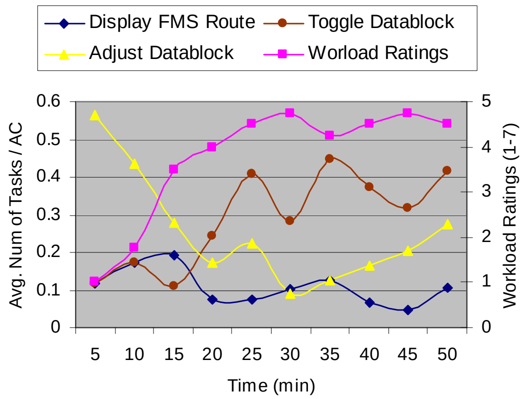


Fig. 4a. Number of “bookkeeping” tasks vs. workload ratings for high traffic scenarios

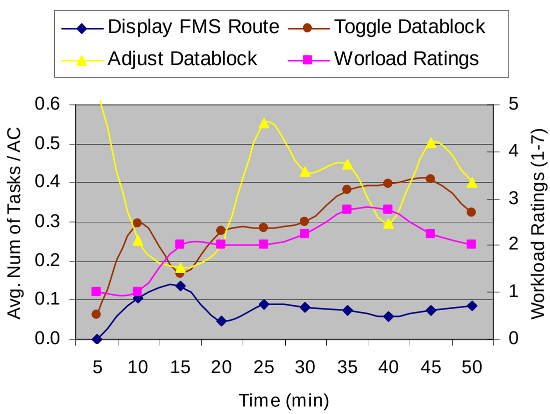


Fig. 4b. Number of “bookkeeping” tasks vs. workload ratings for moderate traffic scenarios

Only the following monitoring/bookkeeping actions that occurred with enough frequency to be analyzed: datablock adjustment, datablock toggle, and FMS route display. Each of these tasks correlates strongly with one or more bookkeeping tasks that controllers engage in to maintain situation awareness. For example, FMS routes

were displayed when controllers wanted to see their routes to resolve potential conflicts or to provide service to the aircraft. Figs. 4a and 4b suggest controllers display routes during the earlier phases of the scenarios when they have more time to plan their actions strategically. In Fig. 4a, the data in the high traffic scenarios suggest that controllers display the FMS routes for a peak of 20% of the aircraft 15 minutes into the scenario and less so thereafter. Fig. 4b shows a similar result for the moderate traffic scenarios, in which controllers display FMS routes at a peak (approx. 15% of the aircraft) at 15 minutes into the scenario. Both data show slight increase in the task frequency during dips in the workload ratings, suggesting that this task is done when workload permits.

One of the interesting results is the frequency of datablock adjustments in the high traffic scenarios. Two main reasons for datablock adjustments were 1) to organize the datablocks so that a plane entering a sector, actively being managed, or a plane ready to be shipped to the next sector each has a particular datablock orientation to remind the controllers of its current status, and 2) to keep them from overlapping on the screen. For these reasons, one would predict that datablock adjustment per aircraft would either stay relatively constant or increase slightly with increased levels of traffic. However in the *high traffic scenarios*, the data suggest that the frequency of this task sharply decreases (from about 60% of the aircraft to 10%) as the workload increases to its peak (around the 30 minute mark) and increases again as the scenario continues (see Fig. 4a). The results support the hypothesis that controllers shed this task when the workload is high, presumably because it is a lower priority task during peak workload. What is interesting, however, is that this pattern of results is not duplicated in the *moderate traffic scenarios*. Fig. 4b shows that although the frequency of datablock adjustment per aircraft dropped to approximately 20% at 15 minutes, the rate climbed back up between 30% and 50% for the rest of the scenario. Combined results suggest that high (but not moderate) workload situations reduce the rate of datablock adjustments, likely because controllers manage their workload by minimizing the frequency of peripheral tasks.

Another interesting finding is from datablock toggle task. In this simulation environment, toggling a datablock almost exclusively had one function – to minimize the datablock after the plane has been handed off to the downstream sector. This was an important task to minimize display clutter but the timing of the event was less critical. Given this understanding, it was interesting to see an “oscillation” pattern in this task that was out of phase with workload data in the *high traffic scenarios*. The result suggests that whenever the workload is at its peak, controllers delayed minimizing the datablock until the workload dipped slightly from its peak (see Fig 4a). In contrast, the oscillating pattern is not present in the *moderate traffic* condition, suggesting that this task was not delayed when the workload was not at its maximum.

Workload vs. Route/Altitude Clearances. The data used for this analysis were collected for a future operational environment that allowed route and altitude trial planning capabilities that could construct conflict-free 4-D trajectories graphically using the trackball. Controllers could then data link the conflict-free paths to the flight deck, bypassing voice clearances altogether. The only situations that required voice commands were when the clearances needed to be delivered right away or if the pilot had not responded to data link clearances. In such cases, controllers verbally assigned

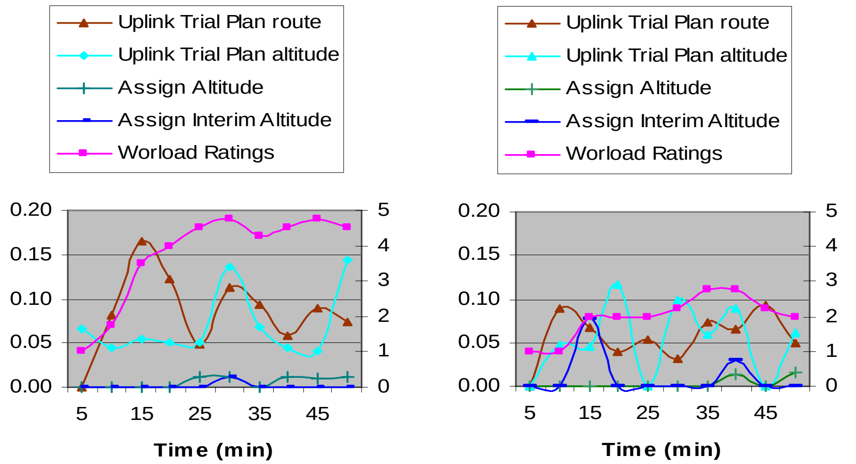


Fig. 5. Number of route and altitude clearances vs. workload ratings for high (*left*) and moderate (*right*) traffic scenarios – y-axes represent the average number of tasks per aircraft on the left and the workload ratings (1-7) on the right side of the graphs

regular and interim/temporary altitudes (shown below) as well as heading and speed changes (not shown). Fig. 5 shows the graphs for the four clearance types described above.

If the controllers minimize their workload in high traffic/workload situations by shedding lower priority and higher workload tasks, one would expect that, given a choice between a lateral and an altitude clearance, they would use the altitude clearance during high workload situation because it requires less work than a lateral route clearance due to an easier interface and less traffic complexity in the vertical dimension. On the other hand, the controllers have commented that they generally try a lateral solution first since they preferred to leave the planes at their preferred altitudes and also want to have the altitude solution available as an “out” maneuver if any last-minute maneuvers are needed. Given these two constraints, one would expect that there would be a greater number of lateral route maneuvers during the low to moderate workload situations and a shift to a greater number of altitude maneuvers during high traffic situations.

Fig. 5 suggests data trends that generally support the above hypothesis but the details of the data are difficult to interpret. As shown in the left graphs in Fig. 5, later route maneuvers in the high traffic scenarios are issued for about 15% of the aircraft at the beginning of the scenario and taper off to about 5% near the end while the altitude maneuvers are flat at about 5% of the aircraft until mid-scenario where they are issued to about 14% of the aircraft, suggesting that the route clearances are used more often during low to moderate workload situations and the altitude clearances are used most often during peak workload situations. In the moderate traffic scenarios, the lateral route clearances have higher frequencies during lower workload situations than during higher workload situations, suggesting again that the route clearances are used more often during these periods. However, altitude clearances seem to be issued periodically throughout the scenario, suggesting that they are not used specifically to

minimize workload in the moderate traffic situations. There also seem to be periodic oscillation patterns present in these data, but more analyses are needed to understand the exact nature of the oscillations.

Workload vs. Handoff-related Tasks. Similar to the traffic load test data, the number of handoffs that a controller accepts from an upstream sector and initiates to a downstream sector is directly related to the number of aircraft in their sector. The average frequency of handoff initiation and acceptance per aircraft therefore shows considerable similarity between the high and moderate traffic/workload situations (see Fig. 6). The similarity between the two graphs suggest that these tasks are important and could not be shed or delayed significantly during maximum workload situations, unlike the bookkeeping tasks and the clearances described in the previous sections.

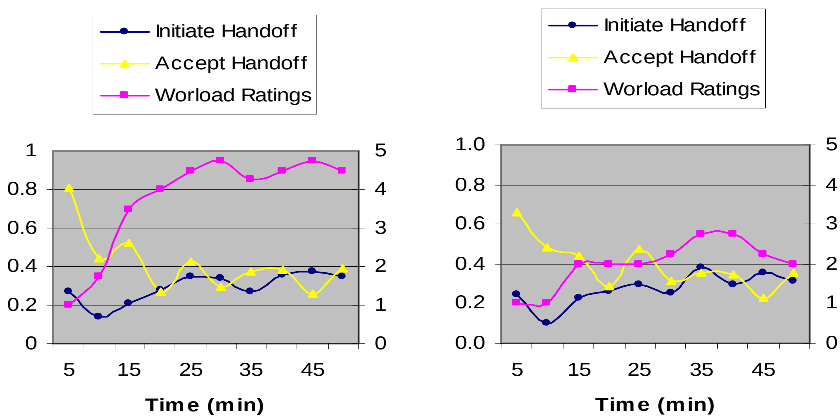


Fig. 6. Number of handoff-related events vs. workload ratings for high *(left)* and moderate *(right)* traffic scenarios – y-axes represent average number of tasks per aircraft on the left and the workload ratings (1-7) on the right side of the graph

4 Conclusion

Based on the examination of how workload affects task distribution, the results suggest a qualitative shift in the types of tasks that controllers perform in low, moderate, and high workload states. During low traffic/workload states, controllers engaged in a significantly higher percentage of “bookkeeping” activities, such as datablock adjustments, than in the high workload states. If the workload reaches the maximum such that controllers need to manage their workload by selectively shedding or delaying tasks, it seems that they shed datablock adjustments and delay toggling/minimizing datablock until they have enough time to attend to this task. None of these patterns emerge in the moderate traffic scenarios, presumably because controllers have enough mental resources to perform all of the tasks.

Examining route vs. altitude clearances, we expected controllers to issue more altitude instead of lateral route clearances during peak workload situations because

altitude clearances generally take less workload. Although the data seem to support this hypothesis, they were not conclusive.

Understanding how workload moderates task load distribution has significant potential for predicting true workload limits. If the pattern of delayed and dropped tasks show better consistency across controllers than the subjective workload ratings themselves, one can look for these patterns to indicate when the controllers are reaching their mental resource limits, which in turn could provide inputs to safety implications and capacity limits.

Acknowledgments. This study heavily leveraged previous simulation data from Distributed Air-Ground Traffic Management (DAG-TM) research, which was funded by Advanced Air Transportation Technologies (AATT) Project. The authors would like to thank members from the NASA Ames Airspace Operations Laboratory (AOL) in providing access to the data as well as continued support in this research. Finally, we would like give special thanks to Natalia Wehrle for her help in the data compilation/analyses.

References

1. Athenes, S., Averty, P., Puechmorel, S., Delahaye, D., Collet, C.: ATC Complexity and Controller Workload: Trying to Bridge the Gap, In: HCI-Aero: International Conference on Human-Computer Interaction in Aeronautics, Cambridge, MA (2002)
2. Hilburn, B.G., Bakker, M.W.P., Pekela, W.D, Parasuraman, R.: The effect of free flight on air traffic controller mental workload, monitoring and system performance. In: Proceedings of the 10th International CEAS Conference on Free Flight pp. 14–1/14–12 (1997)
3. Stein, E.: Air Traffic Controller Workload: An Examination of Workload Probe, DOT/FAA/CT-TN84/24. DOT/FAA Technical Center, Atlantic City, NJ (1985)
4. Kopardekar, P., Magyarits, S.: Measurement and prediction of dynamic density. In: 5th USA/Europe Air Traffic Management R&D Seminar, Budapest, Hungary (2002)
5. Lee, P.U., Prevot, T., Mercer, J., Smith, N., Palmer, E.: Ground-side Perspective on Mixed Operations with Self-separating and Controller-managed Aircraft, In: Proceedings of the 6th USA/Europe Air Traffic Management Research and Development Seminar, Baltimore, MD (2005)
6. Manning, C., Mills, S., Fox, C., Pfleiderer, E., Mogilka, H.: The Relationship Between Air Traffic Control Communication Events and Measures of Controller Taskload and Workload. In: The Fourth International ATM R&D Seminar ATM-2001, Santa Fe, NM (2001)
7. Lee, P.U.: A Non-Linear Relationship between Controller Workload and Traffic Count. In: 49th Annual Meeting of Human Factors and Ergonomics Society, Orlando, FL (2005)
8. Lee, P.U., Prevot, T., Mercer, J., Smith, N., Palmer, E.: Ground-side Perspective on Mixed Operations with Self-separating and Controller-managed Aircraft, In: Proceedings of the 6th USA/Europe Air Traffic Management Research and Development Seminar, Baltimore, MD (2005)

Cognitive, Perceptual, Sensory and Verbal Abilities as Predictors of PDA Text Entry Error and Instructions Across the Lifespan

Hiroe Li and Peter Graf

University of British Columbia, 2136 West Mall, V6T 2G9, Canada
{hiroeli,pgraf}@psych.ubc.ca

Abstract. Sixty-three participants (range from 18 to 85 years of age) completed 4 data entry tasks on an HP iPAQ 5450 via a touch-screen QWERTY keyboard, as well as a battery of neuro-cognitive tests. Entry errors and assistance required by participants were coded into categories. Multiple regression analyses revealed that episodic memory was the strongest predictor for stand-still errors and commission errors, while sensory abilities was the strongest predictor of omission errors. We suggest that raining sessions that familiarize older adults with the functions of specific keys (e.g. Spacebar and Backspace) and structure of the keyboard, complemented with visual or auditory feedback provided by the keyboard as methods to improve text entry accuracy.

Keywords: Human-Computer Interaction, PDAs, aging, text entry errors.

1 Introduction

Personal Digital Assistants (PDAs) are portable and powerful devices that offer many of the functionalities of standard PCs, such as a calendar book, contact list and internet capabilities. In addition, a range of software applications can be added that could potentially be used to assist older adults. For example, mobile computers can be a lifeline for those with memory problems accompanying normal aging, such as serving as reminders to take medication at specific times [15] or function as a way finding systems [11]. It can also provide a portable means of communicating with family, friends and healthcare providers.

However, issues relating to data input have been highlighted as an ergonomics problem that arises from such mobile computing devices [2]. It is documented in the literature that older adults make more text entry errors [4, 8, 9] that are attributed to declining sensory, perceptual, motor and cognitive functioning that accompany healthy aging [7, 14, 17]. Although recent developments have been successful at decreasing target acquisition errors [19, 23, 30] that are hypothesized to arise from poor visual and motor functioning, text entry is rarely perfect. In order to complement the effective use of these text entry developments, we need to support older adults users by developing training procedures that familiarize them with text entry, as well as specify design guidelines of the keyboard that are suited to the capabilities of these users.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 349–358, 2007.

The first objective of the present study was to explore the types of text entry errors made and the assistance required to correct these errors. Three studies have categorized transcription errors (on a normal-sized keyboard) as substitution, intrusion, omission, and transposition [13, 21, 22]. These errors have been interpreted as reflecting perceptual confusion, failure to monitor the phrase sequence, and faulty assignment of movement specification [20]. On a handheld computer, Wright and colleagues [31] noticed that the most common error was to omit spacing between words among older adults. This was thought to reflect unfamiliarity with the keyboards. Errors resulting from wrong pressure and tapping the wrong letter were also noted. These errors have been attributed to arise from difficulties in visually discriminating among certain letters on the touch-screen keyboard. However, no empirical validation was provided for these speculations. In addition, no study has reported what types of instructions would be required to assist with text entry.

To explore the underlying sensory abilities and cognitive processes that may explain the occurrence of these errors and instructions, we employed a battery of neuro-cognitive tests to explore the predictive power of sensory-perceptual abilities, perceptual speed, episodic memory and verbal intelligence on the types of text entry errors and help instructions. One study has tied text entry performance to these abilities. Czaja and Sharit [9] demonstrated that visuomotor skills and memory predicted typing errors above and beyond computer experience. However, only the total number of errors was used as the dependent variable. Thus, the second objective of our present study will be to establish an empirical link between these abilities and different types of entry errors and instructions.

2 The Present Study

2.1 Methods

2.1.1 Participants

Sixty-three healthy community-living adults (43 women; 20 men) participated in this study. Participants ranged from18 to 85 years of age, with an average age of 49.8 (*SD* = 18.8) years.

2.1.2 Apparatus

A Hewlett-Packard iPAQ 5450 handheld computer running on Microsoft Windows® for Pocket PC 2002 was used for this experiment. The screen measured 2.26 inches wide by 3.02 inches tall and had a transflective LCD (64, 000 colors). A touch screen QWERTY keyboard was available for data entry that occupied approximately 1/3 of the area of the screen. The alphabets and number keys on the keyboard measured 3mm by 4mm in size.

2.1.3 Cognitive Tests

We employed a battery of standardized neuro-cognitive tests for assessing sensory abilities, perceptual speed, episodic memory and verbal intelligence [listed in Table 1]. Each selected test is the golden standard for assessing the abilities described and has been shown to have high validity and reliability. For more detailed descriptions of each test, refer to [12, 27].

2.1.4 Procedures

Each participant entered a series of short phrases in the PDA using a stylus via a QWERTY keyboard while sitting next to the experimenter. Information required for text entry tasks remained in view for the duration of each task. Button-presses and stylus interactions with the PDA, as well as all communications between the participant and the experimenter, was recorded by a Hitachi DZ-MV380A digital video camera.

After completing the PDA tasks, the experimenter administered the battery of neuro-cognitive tests listed in Table 1.

All text entry errors as well as help-requests and instructions were transcribed and coded from the video records. Errors were coded according to the criteria described by Wobbrock and Myers [29]. Each error was grouped into one of the following 7 categories:

DOUBLE: creating an unnecessary duplicate character after a target character
WORD: omitting characters within words or whole words in the designated phrase
LANDING: landing on keys adjacent to the target key
LAYER: unnecessarily switching between the upper/lower case layer of the keyboard (e.g. by pressing SHIFT or CAP lock)
CASE: entering a target character in the wrong case (i.e. creating a lowercase letter when it is supposed to be in uppercase, or vice versa)
IRRELEVANT: creating irrelevant characters in relation to the designated phrase.
NEEDED: deleting characters that are relevant to the designated phrase

Based on the video records, we also grouped all participant requests for help/information, confirmation or clarification into the following 9 categories:

LOCATE: showing participants the location of the keys/buttons
FUNCTION: explaining the use of a particular key
TYPO: pointing out to participants that there is either a missing character in the word or an extra unneeded letter in the word
ERASE: instructing Ss to erase/delete unwanted info
CONFIRM: experimenter validates that a key is correct when the participant asks
RULES_W: explaining rules of using a word processor
RULES_K: explaining rules of using a keyboard
KEY: instructing participants to press a specific key/button
REMIND: reminding participants the designated phrase to be entered

2.2 Results

The data was screened for outliers, defined as any value that was more than 3 standard deviations away from the average. Each outlier was replaced with a value 3 standard deviations either above or below the mean. For ease of visual display, the results are shown in 3 groups, each with an equal number of participants: younger adults (18-39 years), middle-aged adults (40-60 years), and older adults (61-85 years).

Table 1. Performance on the neuro-cognitive tests and results of regression analyses that used participant's age as a predictor

Neuro-cognitive tests	Age Group			R^2 Age	Ability
	Y	M	O		
	M	*M*	*M*		
Sensory abilities					
HSI [5]	23.14	22.40	26.33	.08*	auditory sensitivity
CSI [6]	11.38	11.80	12.71	.04	color discrimination ability
Snellen eye chart [1]	9.05	7.43	7.33	.33*	visual acuity
Perceptual-Motor					
TMT† [25]	22.10	40.01	49.81	.08*	motor speed
DSST† [18]	128.14	156.99	174.52	.36*	perception
Stroop† [12]					attention and inhibition
Word Reading	35.95	36.86	38.62	.02	
Color Naming	48.24	50.94	55.86	.23*	
Incongruent	79.52	91.35	105.95	.20*	
Episodic Memory					
RAVLT⌂ [16]					episodic memory
List A: 1st recall	8.48	8.14	7.00	.11*	
List A: 2nd recall	12.67	12.31	9.67	.17*	
List A: 3rd recall	14.90	14.40	10.95	.22*	
List B recall	8.00	6.12	5.67	.16*	
List A: Short delay	9.86	10.22	6.05	.14*	
List A: Long delay	11.00	9.66	6.10	.24*	
Verbal ability					
NAARTЖ [3, 28]	29.95	22.57	22.00	.13*	vocabulary

* indicates $p < .05$ Ж indicates number of words pronounced incorrectly
† completion time   ⌂ indicates number of words correctly recalled

2.2.1 Neuro-cognitive Tests

A principle components analysis using a Varimax rotation confirmed that each test loaded on its respective factor. Table 1 displays the neuro-cognitive test performance across the 3 age-groups, as well as the age-related effects for each test component. In summary, performance on these tests was comparable to results obtained in previous research [12, 16, 28] indicating that our sample consisted of a group of cognitively healthy adults.

2.2.2 Data Entry Errors

Table 2 shows the summary for each type of text entry errors, as well as R^2 with age. The strongest age-related effect was found for typing words in an incorrect case, with older adults making 3 times more of these errors than younger adults. Older adults were also twice as likely to delete needed characters and to type in duplicate characters (presumably reflecting unsteadiness in applying pressure to the keyboard)

compared to younger adults. Errors that involved landing in adjacent keys and errors that created unneeded characters showed no significant age-related effects. Younger adults made as many landing errors as older adults. It is possible that in fact older adults made Landing errors due to hand unsteadiness while younger adults were less cautious with entering text perfectly because they had the knowledge on how to reverse these errors.

Table 2. Basic descriptive statistics of each of the error types from each age group and the relationship between age and performance on each of the errors

Entry Errors	Age Group			R^2 Age
	Y	M	O	
	M	*M*	*M*	
Case	1.76	3.89	5.48	.23**
Double	.62	1.35	1.55	.15*
Irrelevant	2.19	2.70	3.85	.05
Landing	2.29	.72	2.19	.00
Layer	1.29	1.34	2.84	.09*
Needed	4.52	5.20	8.19	.12*
Word	.49	.62	.84	.07*

* indicates p < .05
** indicates p < .01

Table 3. Basic descriptive statistics of each of the instruction types from each age group and the relationship between age and performance on each of the instructions

Entry Ins	Age Group			R^2 Age
	Y	M	O	
	M	*M*	*M*	
Confirm	.20	1.02	1.97	.26**
Erase	0	.49	1.39	.38**
Function	.19	1.23	2.85	.40**
Key	.38	1.96	5.78	.36**
Locate	.08	1.51	4.54	.49**
Remind	.81	1.44	3.93	.41**
Rules_K	.33	1.05	3.01	.29**
Rules_W	0	.65	2.02	.25**
Typo	.29	.69	1.32	.20**

** indicates p =< .01

To identify the distinctiveness of the data entry error types, a principle components analysis with a Varimax rotation of the factors was conducted. The analysis revealed 3 distinct groupings, interpreted as Standstill (Double, Layer and Case), Commission (Irrelevant and Landing) and Omission (Word and Needed) errors.

2.2.3 Text Entry Instructions

Regression analyses with each instruction type on age revealed that age accounted for up to 50% of the variance in the data. Older adults required a substantially greater number of each type of help than the other participants [Table 3], suggesting that older adults were less familiar with the functions and location of keys. Among the request for location of a key, the Spacebar and Backspace were the most common. In the Rules_K category, the most common explanations given included telling participants that the Caps Lock must be off in order to use the Backspace key. For the Key category, the frequently given phrase is to press the Shift key to switch layers to see the hidden symbols on the number line.

A principle components analysis with a Varimax rotation was conducted to explore the distinctiveness of the types of text entry instructions. The analysis revealed 3 distinct factors, interpreted as Keyboard (Function, Key, Confirm and Locate), Rules (Rules_W and Rules_K) and Phrase (Erase, Typo, and Remind).

Table 4. Cross-correlations between entry errors and entry instructions

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Entry Errors																
1. Case	/															
2. Double	.37**	/														
3. Irrelevant	-.03	.11	/													
4. Landing	.05	-.12	.27*	/												
5. Layer	.35**	.39**	.03	.09	/											
6. Needed	.12	.20	.22	.29*	.17	/										
7. Word	.30*	.20	-.02	.06	.12	.38**	/									
Entry Instructions																
8. Confirm	.29*	.44**	.27*	-.10	.30*	.13	.23	/								
9. Erase	.29*	.44**	.19	.03	.28*	.21	.27*	.50**	/							
10. Function	.40**	.27*	.20	-.02	.30*	.33**	.31*	.44**	.49**	/						
11. Key	.54**	.49**	.20	.02	.62**	.33**	.35**	.59**	.52**	.70**	/					
12. Locate	.38**	.39**	.36**	.05	.34**	.36**	.24	.51**	.53**	.77**	.80**	/				
13. Rules_W	.31*	.36**	.29*	.20	.31*	.39**	.32*	.57**	.67**	.57**	.60**	.68**.	/			
14. Rules_K	.52**	.47**	.14	.05	.69**	.44**	.16	.43**	.43**	.59**	.79**	.62**	.59**	/		
15. Remind	.24	.49**	.30*	.18	.51**	.39**	.22	.36**	.50**	.35**	.60**	.69**	.59**	.62**	/	
16. Typo	.43**	.39**	.15	.01	.20	.36**	.40**	.3***	.50**	.44**	.42**	.47**	.61**	.53**	.42**	/

* indicates $p < .05$

** indicates $p =< .01$

Table 5. Regression of entry errors and instructions on neuro-cognitive components

	Standstill	Commission	Omission	Keyboard	Rules	Phrase
Verbal intelligence	-.11	-.11	.03	-.19	.00	-.16
Sensory abilities	-.19	.05	-.27*	-.22	-.10	-.25*
Perceptual-motor	.19	.07	.23	.03	.12	.33*
Episodic memory	-.41**	-.33*	-.02	-.32*	-.50*	-.16

† each number represents a standardized beta coefficient

* indicates $p < .05$

** indicates $p =< .01$

2.2.4 Relationship Between Text Entry Error and Instructions

All inter-correlations among the types of text entry instructions were significant, and these relationships ranged from moderate to strong [displayed in Table 4]. Very few significant inter-correlations were found among types of text entry errors, suggesting that there may be clusters of mistakes that participants make (i.e. making one type of error are also linked to making more of other types of errors). In particular, individuals that made more Double errors also made more Case and Layer errors. Participants that deleted Needed characters also made an increased number of Word errors, suggesting problems in monitoring text entry progress. Medium-sized correlations were found across text entry instructions and errors, suggesting that individuals making any types of entry errors required external instructions in assisting with task completion. Landing errors were shown to be not significantly related with requiring any extra help. One reason for this is that Landing errors follow a curvilinear trend with age, and a fundamental requirement for computing correlations is that both variables must fit a linear model.

2.2.5 Relationship of Errors and Instruction Categories with Neuro-cognitive Components

Multiple regression analysis was conducted to explore the predictive nature of the neuro-cognitive components on each error and instruction grouping. Better episodic memory was related to making less Standstill and Commission errors, and Keyboard and Rules instructions [Table 5]. This confirmed earlier speculations that monitoring phrase sequence is important in making correct entries. As expected, participants who forgot explanations of instructions given earlier were more likely to require these instructions again.

Poor sensory abilities predicted a greater number of Omission errors, while both worse sensory and perceptual-motor abilities were found to be related to requiring a greater number of Phrase instructions. These instructions were mainly related to pointing out to participants that there was a mistake in the entry and to correct this mistake. Poorer sensory abilities explained the failure to detect mistakes made in the entry. The increase of these instructions is also likely due to slower responses by the participants, whereby then the experimenter had to repeat the instructions several times because it was inferred that the participant did not hear or understand the instructions.

Verbal abilities were not a significant predictor of any categories of error or instruction. Although older adults were performing worse than younger adults on text

entry and required more recovery instructions, aging is associated with improved verbal abilities. Older adults may have been able to use their verbal abilities to compensate for the negative effects of sensory and cognitive declines, thus attenuating the relationship.

3 Discussion

One of the major findings from this study reveals that not all types of errors are committed equally by older adults, and that they require all various types of instructions to assist in undoing these errors. In particular, the most frequent errors involved typing characters in the incorrect case. Poor episodic memory performance is one of the causes for the greater number of Case errors among older adults. These errors are most likely to arise when older adults forget to turn off Caps Lock to type lower case characters. In order to counteract this memory-related deficit, one recommendation would be to alert the user, either using visual or auditory cues, that the Caps Lock button is on when the entered characters are not at the beginning of a sentence.

Another explanation for why older adults make more of these errors may be due to the structure of the keyboard. The organization of layers (i.e. upper and lower case) seen in keyboards is often employed because the available space on the screen is very small on PDAs. The keyboard often requires the use of special function keys to access certain numeric and punctuation characters. It is known that moded styles of interaction can be confusing to users [24, 26] and this effect may be magnified with users who are unfamiliar with this type of arrangement. This is exemplified in the requests for Rules_K instructions by older adults (e.g. turn off Caps Lock to see the lower case characters). It appears that older adults are not familiar with rules associated with function-related keys upper case characters. Explicitly incorporating this type of instructions during the training sessions may ease confusion.

Hand unsteadiness and pressure-related errors were also noted among the participants. One input alternative to address these errors is for keys to be selected by removing the pen from the screen rather than by tapping [19]. However, this method is most likely to slow down typing significantly and is not particularly useful for users that do not have these problems. One suggestion would be to add this option under the input devices and allow the user to activate this option if desired.

One difficulty in entering text on a PDA is that it does not allow simultaneous monitoring of the keyboard and entered text. Thus, it is difficult to detect whether there are missing letters in the entry (Omission errors). For older adults, this is found to occur more frequently and that poorer sensory abilities are likely to magnify this relationship. There is evidence that adding non-speech sounds can enhance the usability of numerical keypads on small computer touch-screens [4] but the auditory feedback provided by the keyboard will depend on the noise levels in the working environment. An alternative would be to provide a visual indication that a character is missed, such as blinking the screen.

In terms of helping instructions, one of the common requests from older adults is to point out the location of the Spacebar and Backspace key. This group of participants was also more likely to ask for it again even after receiving the instructions earlier.

Previous research has demonstrated that older adults are able to learn new skills; they just take longer than younger adults [10]. Developing training sessions that include instructions that specifically point out the location of these 2 keys can prove helpful. Also, enabling the key to show a reminder when the key is pressed (e.g. a bubble showing SPACE for the Spacebar key) can help older adults in learning the keys better after the training session.

As explained in the Introduction, pocket computers can be simple to use and highly informative reminding devices, and as such could offer useful support to older people at work and elsewhere. The findings in this study indicate that declines in sensory and cognitive abilities impede accurate text entry. Suggestions are provided that may reduce text entry errors among older adults. However, the suggestions provided are based on findings obtained from a combination of statistical techniques and observations made during the study. Future usability studies will be needed in order to ascertain whether these guidelines and suggestions can indeed improve text entry for older adults.

References

1. Arai, M., Katsumi, O., Paranhos, F.R.L., De Faria, J.M.L., Hirose, T.: Comparison of Snellen acuity and objective assessment using the spatial frequency sweep PVER. Graefe’s Archive for Clinical and Experimental Ophthalmology 235(7), 442–447 (1997)
2. Baber, C., Stanton, N.A., Johnson, G.I.: From public technology to ubiquitous computing: Implications for ergonomics. Ergonomics 41, 921–926 (1998)
3. Blair, J.R., Spreen, O.: Predicting premorbid IQ: A revision of the national adult reading test. The Clinical Neuropsychologist 3, 129–136 (1989)
4. Brewster, S.A., Cryer, P.G.: Maximizing screen space on mobile computing devices. In: Summary Proceedings CHI’99, pp. 224–225. McGraw Hill, New York (1999)
5. Coren, S., Hakstian, A.R.: The development and cross-validation of a self-report inventory to assess pure-tone threshold hearing sensitivity. Journal of Speech and Hearing Research 35, 921–928 (1992)
6. Coren, S., Hakstian, A.R.: Color vision screening without the use of technical equipment: scale development and cross-validation. Perception and Psychophysics 43, 115–120 (1988)
7. Craik, F.I.M., Salthouse, T.A. (eds.): The handbook of aging and cognition. Erlbaum, Hillsdale (2000)
8. Czaja, S., Hammond, K., Blascovich, J., Swede, H.: Age-related differences in learning to use a text-editing system. Behavior and Information Technology 8, 309–319 (1989)
9. Czaja, S., Sharit, J.: Ability-performance relationships as a function of age and task experience for a data entry task. Journal of Experimental Psychology: Applied 4(4), 332–351 (1998)
10. Fisk, A.D., Rogers, W.A., Giambra, L.M.: Consistent and varied memory/visual search: Is there an interaction between age and response-set effects? Journal of Gerontology: Psychological Sciences 45, 81–87 (1990)
11. Goodman, J., Gray, P., Khammampad, K., Brewster, S.: Using landmarks to support older people in navigation. In: Brewster, S., Dunlop, M.D. (eds.) Mobile Human-Computer Interaction – MobileHCI 2004. LNCS, vol. 3160, pp. 38–48. Springer, Heidelberg (2004)
12. Graf, P., Uttl, B., Tuokko, H.: Color- and picture-word Stroop tests: Performance changes in old age. Journal of Clinical and Experimental Neuropsychology 17(3), 390–415 (1995)

13. Gurdin, J.T.: Error patterns in skilled and novice transcription typing. In: Cooper, W.E. (ed.) Cognitive aspects of skilled typewriting, pp. 121–144. Springer-Verlag, New York (1983)
14. Ketcham, C.J., Stelmach, G.E.: Movement control in the older adult. In: Pew, R.W., Van Hemel, S.B. (eds.) Technology for adaptive aging, pp. 64–92. National Academies Press, Washington DC (2004)
15. Lanzolla, V.: The usability of personal digital assistants as prospective memory aids for medication adherence in younger and older adults. Unpublished master's thesis, North Carolina State University, United States (2004)
16. Lezak, M.D.: Neuropsychological assessment, 3rd edn. Oxford University Press, Oxford (1995)
17. Lindenberger, U., Baltes, P.B.: Sensory functioning and intelligence in old age: A strong connection. Psychology and Aging 9, 339–355 (1994)
18. Matarazzo, J.D., Herman, D.O.: Base rate data for the WAIS-R: test-retest stability and VIQ-PIQ differences. Journal of Clinical Neuropsychology 61, 351–366 (1984)
19. Potter, R., Weldon, L., Shneiderman, B.: Improving the accuracy of touch screens: an experimental evaluation of three strategies. In: Proceedings of the Conference on Human Factors in Computing Systems, pp. 27–32 (May 1988)
20. Salthouse, T.A.: Perceptual, cognitive and motoric aspects of transcription typing. Psychological Bulletin 99, 303–319 (1986)
21. Salthouse, T.A.: Effects of age and skill in typing. Journal of Experimental Psychology: General 113, 345–371 (1984)
22. Salthouse, T.A.: Anticipatory processing in transcription typing. Journal of Applied Psychology 70, 264–271 (1985)
23. Sears, A., Revis, D., Swatski, J., Crittenden, R., Schneiderman, B.: Investigating touchscreen typing: The effect of keyboard size on typing speed. Behavior and Information Technology 12, 17–22 (1991)
24. Smith, C.C., Irby, C., Kimball, R., Verplanck, B., Harslem, E.: Designing the start user interface. Byte, 7, 242–282 (1982)
25. Spreen, O., Strauss, E.: A compendium of neuropsychological tests. Oxford University Press, Oxford (1998)
26. Tesler, L.: The smalltalk environment. Byte, 6, 90–147 (1981)
27. Tuokko, H., Kristjansson, B., Miller, J.: Neuropsychological detection of dementia: An overview of the neuropsychological components of the Canadian Study of Health and Aging. Journal of Clinical and Experimental Neuropsychology 17, 352–373 (1995)
28. Uttl, B.: North American adult reading test: Age norms, reliability, and validity. Journal of Clinical and Experimental Neuropsychology 24(8), 1123–1137 (2002)
29. Wobbrock, J.O., Myers, B.A.: Analyzing the input stream for character-level entry errors in unconstrained text entry evaluations. ACM Transactions on Computer-Human Interaction 13(4), 458–489 (2006)
30. Wobbrock, J.O., Myers, B.A., Kembel, J.A.: EdgeWrite: A stylus-based text entry method designed for high accuracy and stability of motion. In: Proceedings of the 16th annual ACM symposium on User interface software and technology, pp. 61–70 (November 2003)
31. Wright, P., Bartram, C., Rogers, N., Emslie, H., Evans, J., Wilson, B., Belt, S.: Text entry on handheld computers by older users. Ergonomics 43(6), 702–716 (2000)

Time Estimation as a Measure of Mental Workload

Mats Lind and Henning Sundvall

Department of Information Science/HCI, Uppsala University,
Box 513,
SE-751 20 Uppsala, Sweden
mats.lind@dis.uu.se

Abstract. Technical systems of different kinds might differ in the mental demands they put on their users while being equally usable in a more conventional sense. Several methods exist that measure mental workload. However, in the everyday practice of usability evaluations, none of these methods seems to be used. This is probably due to the amount of effort needed to use them. The study of our errors in estimation of durations show that errors increase as a function of the amount of attentional resources being needed for other concurrent tasks. This points towards a simple way of estimating mental workload. By asking people to provide estimates of elapsed time after a task, disruptions in their estimates could indicate the mental workload of the task. We conducted a first study aimed at validating this idea with the NASA TLX method. Results show that errors in time estimates correlate significantly with TLX scores.

Keywords: Mental workload, time estimation, usability.

1 Introduction

Mental workload is most often assessed when complex and safety critical tasks are investigated such as the ones found in air traffic control or in the control of nuclear power plants. However, from a usability point of view mental workload could be an important factor to consider in many other areas. It seems reasonable to speculate that high mental workloads could result in worker health problems as well as affecting the quality of work output in many, and more common, work situations (see e.g. [1]). Nevertheless, most usability evaluations of systems intended for use in more ordinary work situations do not, in our experience at least, employ measures of mental workload. One important reason is that funding for usability activities still often is bordering on inadequate. Therefore usability evaluations need to be time and cost effective and usability professionals, consequently, often settle for the bare minimum in terms of measurements. Furthermore, representative users in these types of usability tests are not usually accustomed to the use of special lab procedures or scales and they sometimes give subjective ratings of general usability that are quite inconsistent with the more objectively gathered measurements. These constraints make existing measures of mental workload difficult or impossible to use (for an excellent overview of such measures, see [2] pp. 301-364).

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 359–365, 2007.

The problem remains, however, and we believe that if a measure could be found that is minimally intrusive and practical enough to use in almost any usability evaluation, a whole new insight into many problems associated with computer use in working life could be found. As a step towards finding such a measure we have looked into the use of errors in time estimation as an indicator of mental workload. Time estimation is easy to administer, a simple “x” on a time line is enough, and has a solid theoretical and empirical foundation in cognitive psychology.

1.1 Time Estimation and Mental Workload

Most of us can produce anecdotal evidence of mental workload affecting our perception of elapsed time. We often literally feel time rushing away from us when we are heavily engaged in trying to meet a deadline for a difficult task whereas the minutes drag on while we are inactively waiting for a delayed airplane. More solid scientific support for this phenomenon can be found in the literature on working memory. Here, the notion of a central executive function with limited capacity is very common and a recent discussion of this particular aspect can be found in [3]. A number of experiments have shown that there is a trade-off between temporal processing of intervals in the range of seconds and simultaneous nontemporal information processing (e.g. [4], [5]). For instance, Brown [4] report interference effects when subjects produced 2 and 5 second intervals from such different task as visual pursuit, visual search and mental arithmetic. On the theoretical side these results are often explained by a so-called “attentional-gate” model [6],[7] where the output of a hypothesized time pulse generator must pass a cognitive gate. This gate is controlled by the amount of attentional resources presently allocated to temporal information processing. When more resources are allocated to such processing, timing estimates are more precise and vice versa.

Recently, longer time periods than a few seconds have also been investigated [8]. In one experiment, subjects were asked to give time estimates of durations as long as 40 seconds. Even for these longer intervals the same pattern of results emerged. When working memory was heavily engaged by memorizing a sequence of letters, the errors in time estimation more than doubled compared to the condition with no working memory load.

In light of this it seems safe to say that the empirical evidence so far point to that disruptions of time estimations are caused by attentional resources being heavily engaged in one or several of other, nontemporal, tasks of a varied nature. Of course, mental workload is a multi-dimensional concept [9]}, but how heavily engaged a hypothesized limited attentional resource is, must in any case be of central importance.

Using time estimation as a measure of mental workload could, perhaps, be regarded as using dual task paradigm. In this case, however, the secondary task is something most of us perform more or less constantly. No training or special instructions are thus needed, at least in principle, although some previous studies (e.g. [8]) have employed a calibration procedure.

An interesting difference between the anecdotal evidence and the formal experiments is that the latter use unsigned error as measure of discrepancy between actual time and perceived time. The anecdotal evidence suggests that the sign of the

discrepancy might be interesting as well. Low memory load could perhaps lead to an overestimation of durations whereas high memory load could lead to an underestimation.

1.2 Usability and Mental Workload

Usability, as defined by ISO [10], is measured in terms of user's effectiveness, efficiency and satisfaction for specified task in a particular context of use. Efficiency is associated with the amount of resources consumed in the process of solving the task. In practice, this usually means how much time a user spends while solving a task, but could easily incorporate also the mental resources spent. In this view, mental workload would be a subscale of efficiency. In this way, it would be quite possible that two computer systems used to solve the same task would results in equal amounts of time spent by the users to perform the task while demanding different amounts of mental resources and, thus, still demonstrate different degrees of efficiency for the task.

2 Method

As pointed out by Jex [9] there are inherent logical and conceptual problems in defining and validating measures of mental workload. He therefore recommends the use of a well-known subjective method for initial validation of a new proposal. Hence we chose the NASA-TLX method [11] as our standard comparison. We also decided to use the main scale of the TLX, the "Total mental workload" scale, mainly because we would use only a limited number of tasks and subjects in this initial evaluation of the duration estimation idea. There would also be a limited time for each subject to learn the TLX method before using it, something that could cast further doubts on the use of the measurements in the TLX subscales.

2.1 Subjects

Thirteen persons, 7 male and 6 female, aged between 19 and 31 and with a mean age of 24.6 years, participated voluntarily in the study. They all received a small compensation for participating.

2.2 Tasks

In a simple pilot test with three subjects we investigated 30 candidate tasks. Our aim was to find a subset of tasks that gave a homogenous range of scores on the TLX main scale and that could be solved within the course of one hour. Nine tasks were selected, two as simple warm-up tasks and seven as the main tasks. The selected tasks included finding specific pieces of information on websites, solving computer based puzzles and solving a simple sudoku game. The seven main tasks were administered in random order to each subject.

2.3 Apparatus

All tasks were performed on a PC running Windows XP and was equipped with a 17” TFT monitor. The PC was connected to the Internet. No time related information was visible on the monitor and all subjects were asked to remove any time keeping device they carried before the session began. Each session was recorded using a screen capture program and timings of real elapsed time were later obtained from these recordings. The subjects’ duration judgments were performed by drawing a line on a time scale printed on paper.

2.4 Procedure

When a subject came to the lab s/he was first instructed in the use of the TLX method for about 5 to 10 minutes. Both oral and written instructions were used for this tutoring. The instructions explicitly told the subjects that they would be asked to use the TLX forms and to estimate elapsed time after each completed task. Elapsed time was always estimated before the TLX form was filled in.

3 Results

Interestingly, several subjects showed underestimation of duration for the tasks with a high TLX score. Because of this we decided to use signed errors as our dependent variable for the duration estimations. We simply calculated the following score for each trial and each subject:

$$\text{Duration error score} = (\text{judged duration} - \text{actual duration}) / \text{actual duration} \tag{1}$$

There was a large variation in these scores, both within and between subjects, ranging from –0.4 to 7.6. Values above 4 were found in two persons only. The TLX scores also showed a large variation. They ranged from 5 to 85 with a clear majority of subjects producing values between 20 and 70. The mean TLX values over all subjects for the seven tasks were [22 33 39 52 60 64 69] indicating a reasonably well-balanced set of tasks in terms of mental load.

The main analysis was the correlation for each subject between the TLX scores and the duration error score. Each such correlation is, of course, only based on seven pairs of values making the correlations error prone. However, the correlations were surprisingly stable. With the exception of two subjects who showed correlations close to zero (0.08 and –0.18), the values ranged between –0.67 and –0.91. For the statistical analysis all thirteen correlations were transformed by Fishers’ r to Z procedure [20]. These values were then tested by means of a t-test to see if they were significantly different from zero. They were ($t=-7.25$, $N=13$, $p< 0.00001$). The mean value of these transformed values was –0.9682 which corresponds to $r = -0.75$. A 95% confidence interval around this mean was also calculated. It was [-1.26 –0.68] corresponding to r=[-0.85 –0.59]. The two scatter plots in Figure 1 illustrate the correlation results.

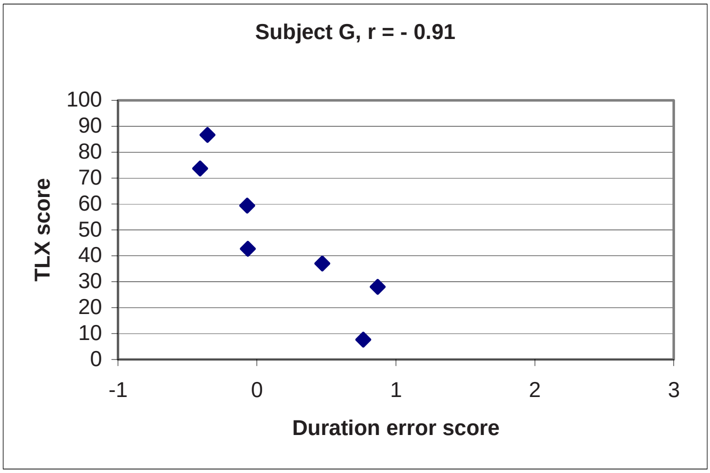


Fig. 1. Scatter plots of duration error scores versus TLX scores for two subjects whose results exhibited the highest and the lowest measured correlation, respectively (not including the two subjects who exhibited correlations close to zero)

The fact these correlations are negative is perhaps surprising but it reflects that the duration of tasks with a low TLX value are systematically more overestimated than duration of tasks with a high TLX value.

4 Discussion

This is of course only a first validation of the basic idea and there are a number of issues that need to be addressed. First of all, there exists a co-variation in our data

between actual task duration and TLX score for most subjects. According to previous results, for instance experiment 3 in [8], errors increase as durations increase regardless of load. However, we got the opposite result; duration estimates are closer to actual duration when task durations are longer. This apparent contradiction could perhaps be explained by the co-variation between mental load, as measured by TLX, and task duration. Larger loads also increase the error in duration estimations, so, given that the sign of these error are opposite, this could result in the errors averaging out for longer durations/higher loads. Since previous studies only report the unsigned error this must remain a speculation until further evidence is gathered. In any case it points to the necessity of conducting more studies where the differential effects of task duration and mental load can be studied.

Another result that needs to be further investigated is the large between subject variation we found in duration estimates. In most previous experiments subjects have been calibrated, for instance by being shown a regularly flashing light before each trial. Doing this would ruin the whole idea of a simple measure of mental workload to use in everyday usability studies. Of course, within subjects designs where the same subject does the same task with two or more interfaces are not affected by this problem since each subject is only compared to herself. Also the possibility of administering a simple calibration task to each subject before the evaluation proper begins could solve the problem, not by calibrating the subject, but by calibrating the measure. Assuming, of course, that each subject shows the same pattern of results over time without any major drift.

Given the small number of subjects and tasks in this study, more studies are needed just to validate the basic results. However, the results we obtained are very promising and it seems to be a worthwhile effort to continue the investigation into the use of time estimation as a measure of mental workload.

In summary it seems appropriate to use the five criteria on a useful measure of mental workload put up by Jex [9]. A good measure of mental workload should be: Relevant, Sensitive, Concordant, Reliable and Convenient.

- The relevance criterion seems fulfilled. The psychological evidence point directly to disruptions in duration estimation as being caused by attentional load.
- Whether or not the measure is sensitive remains to be investigated. Our results seem to indicate that there is at least hope and the procedure can definitely be improved upon.
- The concordance criterion means that the measure should reflect “ubiquitous trends in target population” ([9], p.13). Again, the psychological studies provides hope but more studies are necessary in more applied settings and with more representative samples than the usual university students.
- To find out how reliable this measure is should be a fairly straightforward task and is one of the objectives of planned, forthcoming studies.
- That the measure it is convenient, that it is easy to learn and administer, is portable for use in field trials and evaluations and involves a low cost seems clear already at this point.

References

1. Arnetz, B.B.: Techno-stress: a prospective psychophysiological study of the impact of a controlled stress-reduction program in advanced telecommunication systems design work. J. Occup. Environ Med. 38, 53–65 (1996)
2. Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C., Jenkins, D.P.: Human Facotrs Methods, A Practical Guide for Engineering and Design. Ashgate Publishing Limited, Aldershot (2005)
3. Baddeley, A.D.: Is working memory still working? Am Psychol 56, 851–864 (2001)
4. Brown, S.W.: Attentional resources in timing: interference effects in concurrent temporal and nontemporal working memory tasks. Percept Psychophys 59, 1118–1140 (1997)
5. Fortin, C., Breton, R.: Temporal interval production and processing in working memory. Percept Psychophys 57, 203–215 (1995)
6. Zakay, D., Block, R.A.: Temporal cognition. Current Directions in Psychological Science 6, 12–16 (1997)
7. Zakay, D.: Gating or switching? Gating is a better model of prospective timing (a response to 'switching or gating?' by Lejeune)(1). Behav Processes 52, 63–69 (2000)
8. Venneri, A., Pestell, S., Nichelli, P.: A preliminary study of the cognitive mechanisms supporting time estimation. Percept Mot. Skills 96, 1093–1106 (2003)
9. Jex, H.R.: Measuring mental workload: Problems, progress, and promises. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, vol. 52, Elsevier science publishers, Amsterdam (1988)
10. ISO: ISO 9241-11:1998 Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability. ISO (1998)
11. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, vol. 52, Elsevier science publishers, Amsterdam (1988)
12. Hays, W.L.: Statistics for the social sciences. Holt, Rinehart and Winston, New York (1973)

How Does Distraction Task Influence the Interaction of Working Memory and Long-Term Memory?

Ye Liu and Xiaolan Fu*

State Key Laboratory of Brain and Cognitive Science,
Institute of Psychology, Chinese Academy of Sciences,
Beijing 100101, China
{liuye,fuxl}@psych.ac.cn

Abstract. The present study addressed the influence of distraction task on the interaction of working memory and long-term memory by using available long-term memory tasks with or without distraction task. The results showed that: (a) Distraction task had significant effect on the availability of LTM facilitated by prior attention-driven processing in WM, and (b) the pattern of semantic priming effects observed was reversed between the condition with and without distraction task. These findings support the hypothesis that the semantic activation is implicit automatic process, and less attention resource focused on the process will benefit the semantic activation of LTM.

Keywords: working memory, availability of long-term memory, semantic activation, semantic priming.

1 Introduction

The theoretical concept of working memory (WM) assumes that a limited attentional capacity system, which temporarily maintains and stores information, supports human thought processes by providing an interface between perception and long-term memory [1] [2]. However, relatively small capacity of WM fails to explain complex cognitive activities such as language comprehension, and the greatly expanded working memory capacity of experts and skilled performers [3]. Alternative conceptualizations of WM have been proposed that there should be active long-term memory (LTM) elements in WM. And the active LTM elements are temporarily available for processing, but not in the current focus of attention. These alternative models postulate that a large subset of information or semantic related knowledge in LTM is activated due to prior attention-driven WM processes or well-learned knowledge structures. They also assumed both attention-driven WM processes and automatic LTM activation effectively define capacity limits that constrain complex processing activities [10].

---

* Corresponding author.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 366–374, 2007.

1.1 Automatic LTM Activation in WM Models

Several researchers have proposed alternative WM models emphasizing the role of automatic LTM activation in cognitive activities.

Ericsson and Kintsch [3] have proposed that working memory should include two components. One is the temporary storage of information that they refer to as short-term working memory (ST-WM). Another is based on skilled use of storage in long-term memory (LTM), and they refer to as long-term working memory (LT-WM). Information in LT-WM is stored in stable form, and reliable access to it may be maintained only temporarily by means of retrieval cues in ST-WM. Hence LT-WM is distinguished from ST-WM by the durability of the storage and the need for sufficient retrieval cues. Well-learned knowledge structures and acquired memory skills enables individuals to use LTM as an efficient extension of ST-WM in particular domains and activities after sufficient practice and training. Therefore, LT-WM reflects domain-specific skills of automatically accessing related LTM elements.

Oberauer proposed a model distinguishing three states of representations in WM: the activated part of LTM, a capacity limited region of direct access, and a focus of attention [5]. Oberauer proposed that only one chunk of information was assumed to be directly in the focus of attention at any time. However, a limited number of additional chunks were assumed to be in a state of direct access. Beyond information in a direct access state that based on attention-driven WM, there was a segment of LTM that held some degree of accessibility, but its access would depend on prior attention-driven process.

Woltz and Was further proposed that the content and complexity of attention-driven processing in WM determine the subsequent availability of semantically related elements in LTM [10]. They developed available long-term memory task (ALTM task) to test their hypotheses. In ALTM tasks, semantically mediated priming effects are taken to indicate temporary increases in the availability of LTM. They found a close link between the amount and type of attention-driven processing in WM and the resulting accessibility of semantically related memory structures. Their research also suggested ALTM mediated the relationships of both WM and background knowledge with listening comprehension [9].

According to the previous researches, the interaction of WM and LTM is the foundation of many cognitive activities. Automatic LTM activation facilitated by prior attention-driven processing in WM would influence the concurrent cognitive process.

1.2 The Effects of Additional Mental Load

Working memory is commonly described as a system for simultaneous storage and processing of information, and the two functions share limited common cognitive resource [5]. Evidence from neuropsychological studies showed that working memory, executive control, and focusing of semantic retrieval share a common neural substrate in the prefrontal cortex and are functionally linked [4]. However, dual task studies provide evidence against the resource-sharing hypothesis. A study on visual search suggested that reducing reliance on executive control processes and increasing reliance on rapid automatic processes could improve the efficiency of some difficult searches [8].

Another study on attentional blink also showed additional task load had beneficial effect on the attentional blink, and improved the participants' performance [6].

Since semantic activation produced by prior attention-driven processing is automatic and implicit, distraction tasks that occupy the cognitive resource should have beneficial effect on the availability of LTM. ALTM tasks with distraction task or without distraction task were used in the present study. Our hypotheses were that availability of LTM facilitated by prior attention-driven processing in WM changed with mental load, and additional task load should facilitate availability of LTM.

1.3 ALTM Task

The ALTM task has four trial components (See Fig. 1 for example) [9] [10]. The first component is a memory load that consists of a set of words presented one at a time for eventual recall. Within each memory load, some of the words (usually half) belong to one semantic category, and the remaining words belong to another category. Woltz and Was assumed that the memory load engaged participants' attention-driven WM processes for active rehearsal.

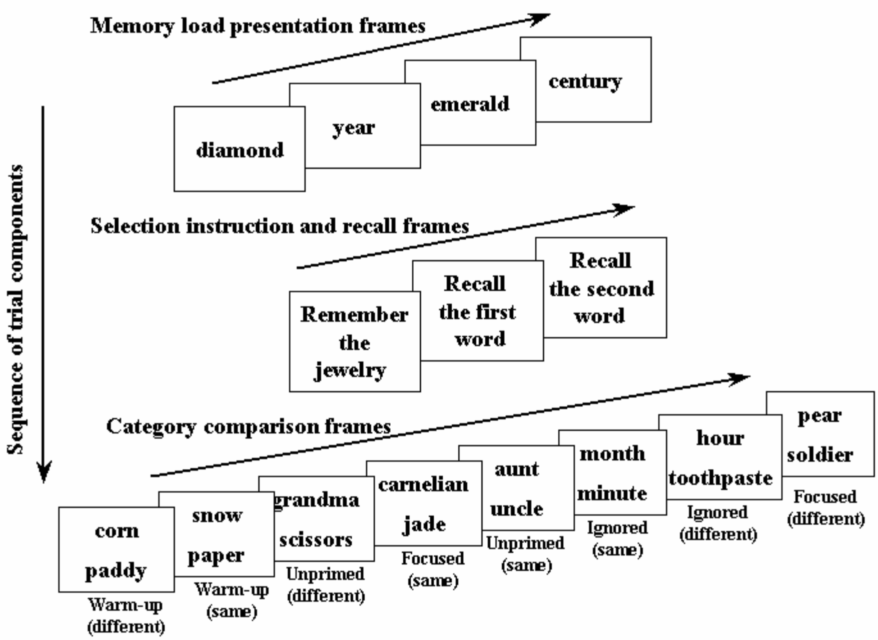


Fig. 1. Example trial of the ALTM task proposed by Woltz and Was [10] used in the present study

The second trial component is the selection instruction. An instruction to remember the words from just one of the categories is presented to participants. This component engages attention-driven category identification processes for one category, and

subsequent rehearsal processes also only focus on the target category exemplars. Therefore, more attentional resources are presumably devoted to the target category.

In the present study, there were two kinds of instruction. One conducted participant to remember the words in the presented category. In this condition, the category was focused target category. The other instruction conducted participant to remember the words that were not in the presented category. In this condition, not target category name but ignored category name was presented to participants.

The third trial component requires participants to recall the words that were remembered. This manipulation verified that individuals were adequately performing the task demands represented in the first two components.

The final type of trial component consists of same-different category membership comparisons that assess semantic priming or availability of the categories represented in the memory load. Within each trial, there were new exemplars from three categories: the focused category in the memory load (i.e., the category that was identified for eventual recall), the ignored category in the memory load, and an unprimed category that was not represented in the memory load. Increased response speed for comparisons representing the two memory load categories (focused and ignored categories), relative to the unprimed category indicated the availability of LTM.

Using ALTM task, the present study investigated the influence of distraction task on the availability of LTM facilitated by prior attention-driven processing in WM.

2 Method

2.1 Participants

Seventy-six undergraduates with a mean age of 22 years (range, 17-25) participated in the experiment (42 men and 34 women) in return for monetary payment. All of them were native speakers of Chinese.

All participants were divided into four groups equally. Two groups were asked to do distraction task before category comparison, while the other two groups were not asked to do. One group with distraction task was instructed with the category identification to be remembered in selection instruction, and another one with distraction task was instructed with the category identification to be ignored. As for the two groups without distraction task, it was the same to the groups with distraction task.

2.2 Apparatus

The participants performed the experimental task on Lenovo microcomputers with standard keyboards. The experiment was programmed with E-Prime software [7].

2.3 Materials

Most of the stimuli used in the experiment were Chinese words with two characters, and a few of them were words with three characters or single character. The semantic categories and exemplars were obtained primarily from the research of Woltz and Was [10], but some of them were revised because of the cultural difference.

All of the categories were organized in 24 sets, with each set containing three categories. For each participant, one category from each set was assigned to be the focused category in the memory load, one was assigned to be the ignored category in the memory load, and the remaining one represented a category not found in the memory load. Six versions of the experiment were created that represented a complete counterbalancing of triplet category assignment to priming condition (focused, ignored, and unprimed). Therefore, each comparison item had ever been under each priming condition twice. In each group, about three participants performed one of the six counterbalanced versions.

For all the positive match category comparison items, there were five kinds of comparison items: (1) belonged to the focused category with category name, (2) belonged to the focused category without category name, (3) belonged to the ignored category with category name, (4) belonged to the ignored category without category name, (5) belonged to the unprimed category. Unprimed condition was a baseline compared with other four conditions.

2.4 Procedure

For the group without distraction task, the experimental task consisted of four sequential components: memory load presentation, selection instruction, memory load recall, and category comparison frames. Each of 24 trials contained the four components in the described order. The sequence of the 24 trials was randomized, and there were 30 seconds interval for participants to have a rest between trials.

Each trial began with the instruction to read words. Then, four words were presented on the display sequently in random order (e.g., diamond, emerald, year, and century). Each word set was preceded by a fixation displayed for 750 ms in the location of the words (center of screen) and then a blank screen for 1 sec. Each word was displayed for 1,500 ms, followed by a blank screen for 500 mss.

Following was an instruction frame that directed participant to remember only two of the four terms in memory load. The instruction to half of participants always named the category to be remembered, (e.g., Remember words that are jewelry). And the instruction to another half of participants always named the category to be ignored, (e.g., Remember words that are not unit of time). The participants could take as long as needed to identify and rehearse the two target exemplars (e.g., diamond and emerald) in the memory load. They were instructed to press the space bar when ready to recall the words.

With the selection instruction frame disappearing, the participants were prompted to recall the two words held in memory with typing the spelling in Pinyin of each Chinese word they were recalling.

Following the second recall frame, there was the instruction to compare words. The participants were instructed to rest their forefingers on the F and J keys, and decide whether the two exemplars in each comparison came from the same category (F response) or different categories (J response). This instruction was self-paced to allow the participants to prepare for the comparison frames.

Each comparison frame began with two asterisks presented for 500 ms, one on top of the other at the location where the two stimulus words would appear. The stimuli remained on the screen until participant responded by pressing either the F or the J

key. There was a total of eight category comparison frames in each trial. The first two frames were warm-ups that contained words unrelated to the contents of the memory load and the unprimed category in the stimulus set. The remaining six frames were presented in random order for each participant. They consisted of positive match frames and negative match frames with the three categories of the stimulus set (focused, ignored, and unprimed).

For the group with distraction task, the experimental task consisted of five sequential components: memory load presentation, selection instruction, memory load recall, distraction task and category comparison frames. Each of 24 trials contained the five components in the same order as the group without the distraction task, except there was a distraction task before category comparison frames. The sequence of the 24 trials was randomized. The distraction task was an arithmetic problem with continuously subtracting a digit (3, 4, 5, and 7 randomly), and lasted 3 minutes.

3 Results

3.1 Accuracy of Recalling and Distraction Task

All of the participants selected and recalled the focused category words from the initial memory loads with high accuracy (100%). This result showed all participants could correctly recall the exemplars that the instruction required to be remembered. For the group with distraction task, the average accuracy rate of arithmetic problem was 76.43%. It suggested the distraction task was much more difficult than memory load task.

3.2 Accuracy of Category Comparisons

The average accuracy rate of positive match condition and negative match condition was 93.60% and 97.70%, respectively. The Pearson's correlation between response time and accuracy rate showed there was no significant speed accuracy trade-off effect in the present study, $r = 0.297$, $p > 0.05$.

3.3 Response Time for Positive Match Category Comparisons

The response time for the positive match category comparisons was showed in Fig. 2.

With items as a random factor, a repeated measure ANOVA was conducted. Priming type (focused or ignored), category identification (category name displayed or not), and distraction condition were three repeated measures. The results showed that the main effect of distraction condition was significant, F_2 (1, 71) = 76.17, $p <$ 0.001, $\eta^2 = 0.52$. The three-way interaction of priming type by category identification by distraction condition was significant, F_2 (1, 71) = 78.79, $p < 0.001$, $\eta^2 = 0.53$. Other main effects and interactions were not significant, $F_2s < 2.5$, $p > 0.05$. The analysis of simple effect showed the role of distraction on hastening response time of comparison under the condition of ignored category and displaying the category name was significant, F_2 (1, 71) = 65.66, $p < 0.001$, and the role of distraction on hastening response time of comparison under the condition of focused category and not displaying the category name was also significant, F_2 (1, 71) = 77.85, $p < 0.001$.

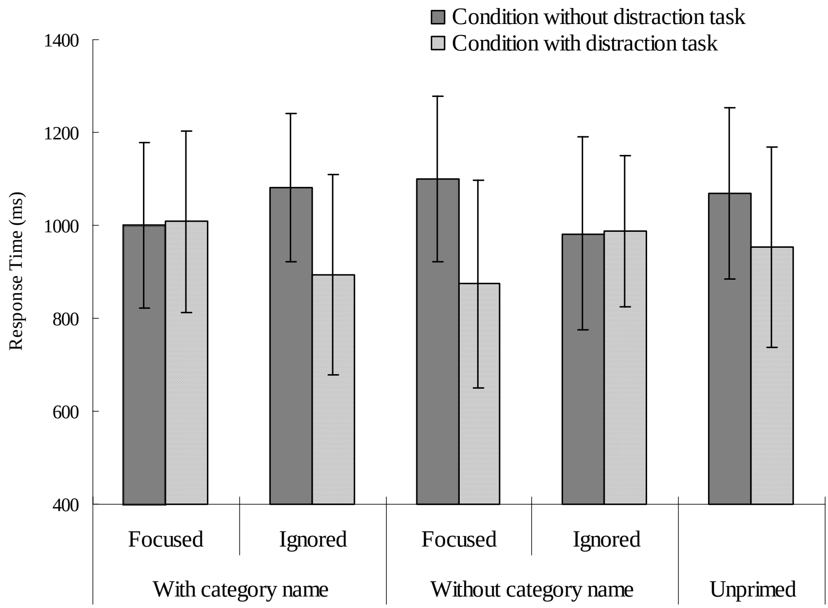


Fig. 2. Mean response time for positive match category comparisons

Because the significant three-way interaction was complicated, Paired-Samples T Tests between four experimental condition and their baselines (unprimed condition) were also performed to explore the reason of interaction.

The results of Paired-Samples T Tests showed: (a) Under the condition without distraction task, if the comparison items belonged to focused category with category name, the participants were significantly faster than the corresponding baseline, $t = -3.01$, $p < 0.01$; (b) under the condition without distraction task, if the comparison items belonged to ignored category without category name, the participants were also significantly faster than the corresponding baseline, $t = -6.09$, $p < 0.001$; (c) under the condition with distraction task, if the comparison items belonged to focused category without category name, the participants were significantly faster than the corresponding baseline, $t = -4.34$, $p < 0.001$; (d) if the comparison items belonged to ignored category with category name, the participants were significantly faster than the corresponding baseline, $t = -6.09$, $p < 0.001$; and (e) other Paired-Samples T Tests were not significant. The results showed that distraction task had significant influence on the priming effects of ALTM task.

4 Discussions

Evidence from the experiment demonstrated distraction task had significant influence on the interaction of working memory and long-term memory. These findings support the hypothesis that the semantic activation is implicit automatic process, and fewer attention resources focus on the process will benefit the semantic activation of LTM.

4.1 Semantic Activation

The results of ALTM task without distraction task showed when the comparison items belonged to the focused category with category name, the responses were significantly faster than the comparison items belonging to the unprimed category. This result suggested there was a significant semantic priming effect, and rehearsal and category identification could facilitate semantic activation of target category. Wshen the comparison items belonged to ignored category without their category name, the responses were also significantly faster than the comparisons belonging to the unprimed category. This result suggested even minimal processing and rehearsal in WM (participants just read the two words belonging to ignored category in memory load presentation) produced significant priming effect. These results were consistent with the found of Woltz and Was [10].

However, the results of ALTM task without distraction task showed that neither the condition of ignored category with category name nor the condition of focused category without category name produced significant priming effect. These phenomena could due to the complex switch the words to be recalled from ignored category name to focused category name. Under these two conditions, much more cognitive resource was involved in the switching process, and inhibited the automatic semantic activation.

4.2 The Role of Distraction Task

The results of the present study showed distraction task had significant influence on the interaction of working memory and long-term memory. The role of distraction task on the availability of LTM facilitated by prior attention-driven processing in WM was observed by ALTM task.

The pattern of semantic priming effects observed was reversed between the condition with and without distraction task. The results of ALTM task c showed when the comparison items belonged to focused category and their category name was not presented, the responses were significantly faster than the comparison items belonged to the unprimed category. And the comparison items belonged to the ignored category and their category name was presented, the responses were also significantly faster than the comparisons belonged to the unprimed category. These results accorded with other researches about the effects of additional mental load [6] [8]. It suggested that distraction task occupied the cognitive resource, and increased the mental load before category comparison. Under these two conditions, there were no more resources to control the category switch. So, rapid automatic process was free from executive control, and it improved the category switch, and produced significant semantic priming effect.

However, the other two conditions presented didn’t get significant priming effect. These phenomena might be due to decay of the activation with time lasting. Although, the semantically mediated priming effects were relatively long lasting, 3 minutes delay set in the present study was too long to maintain the activation.

In conclusion, the findings of the present study suggest the semantic activation facilitated by prior attention-driven processing in WM is implicit automatic process, and less attention resource focus on the process will benefit the semantic activation of LTM.

Acknowledgements. This research was supported by grants from 973 Program of Chinese Ministry of Science and Technology (#2006CB303101), and the National Natural Science Foundation of China (#60433030).

References

1. Baddeley, A.: The episodic buffer: a new component of working memory? Trends in Cognitive Sciences 4, 417–423 (2000)
2. Baddeley, A.: Working memory: Looking back and looking forward. Nature Reviews: Neuroscience 4, 829–839 (2003)
3. Ericsson, K.A., Kintsch, W.: Long-term working memory. Psychological Review 102, 211–245 (1995)
4. Kiefer, M., Ahlegian, M., Spitzer, M.: Working memory capacity, indirect semantic priming, and Stroop interference: Pattern of interindividual prefrontal performance differences in healthy volunteers. Neuropsychology 19, 332–344 (2005)
5. Oberauer, K.: Access to information in working memory: Exploring the focus of attention. Journal of Experimental Psychology: Learning, Memory, & Cognition 28, 411–421 (2002)
6. Olivers, C.N.L., Nieuwenhuis, S.: The beneficial effects of additional task load, positive affect, and instruction on the attentional blink. Journal of Experimental Psychology: Human Perception and Performance 32, 364–379 (2006)
7. Schneider, W., Eschman, A., Zuccolotto, A.: E-Prime User's Guide (Version 1.1). Pittsburgh: Psychology Software Tools (2002)
8. Smilek, D., Enns, J.T., Eastwood, J.D., Merikle, P.M.: Relax! Cognitive strategy influences visual search. Visual Cognition 14, 543–564 (2006)
9. Was, C.A., Woltz, D.J.: Reexamining the relationship between working memory and comprehension: The role of available long-term memory. Journal of Memory and Language (in press)
10. Woltz, D.J., Was, C.A.: Availability of related long-term memory during and after attention focus in working memory. Memory & Cognition 34, 668–684 (2006)

Sequential Analyses of Error Rate: A Theoretical View

Ronald Mellado Miller and Richard J. Sauque

1Brigham Young University – Hawaii,
55-220 Kulanui Street #1896, Laie,
HI 96762 USA
millerr@byuh.edu

Abstract. Though error rate is a ubiquitous measure of human performance, as typically measured in terms of overall error rate or percentage, there are a number of predictive variables lost by summing or averaging the errors made. In this paper, we present a sequential analysis of error rate, where the pattern of errors is analyzed. By examining such concepts as the number of transitions from incorrect responses (I) to correct responses (C) or IC transitions as well a concept called I-length, which refers to the number of incorrect responses followed by a correct response, valid ordinal predictions of persistence in the face of continuous failure can be made. This paper develops this theoretical construct in the hopes that utilizing such data will facilitate the analysis and predictive quality of error rate data.

Keywords: error rate, sequential, performance, percentage, extinction, persistence.

1 Introduction

If a person had bought a high quality new car and that car had worked flawlessly for a year, but, one morning, simply stopped functioning, how long would that person persist in trying to start that car? It had always worked with no difficulties over an extended period of time, its reliability had been shown to be high, etc. Under such circumstances, most people would try a few times and then realize they needed to take the car to the mechanic. On the other hand, think of the car that person might have used as a student during their undergraduate days. It was probably unreliable and may not have started each time—it also may have had other quirks that made it difficult to use. With persistence and the right amount of luck, the car would start. If that car stopped working that morning and, unbeknownst to the driver, it would never start again, how long would that driver continue to try to start their car? It had been unreliable, so it was clear that more effort would be needed to be sure the car really was inoperable. In this situation, the driver might attempt many times to start the car, and certainly many more times than did the driver of the high quality car.

The difference in persistence between the high quality and low quality cars exemplify at least two important points. The first is that that we may spend more time working with a historically unreliable system than with one which has given us better

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 375–378, 2007.

results in the past. This argues against most rational explanations of behavior because we are expending the least amount of effort on the system that has given us the most success in the past. However, when we look at it as a subset of pattern learning, in that the respective drivers had learned what patterns of behavior for a particular system lead to success, then the actions make sense. The other point is that error rate alone does not tell us the nature of the problem. The unreliable car had a higher error rate, in that it started less often than the reliable car, yet it produced the most persistence in the face of failure. For many human-computer interaction studies, the analysis of error rates plays a key role in evaluating the utility and functionality of processes, components, and users. While the concept of utilizing error rates to determine such functionality is clear, as, generally, lower error rates are desirable and processes are designed to monitor and minimize errors, for the end-user, error rate as typically measured becomes problematic. This paper seeks to establish a theoretical model of a sequential analysis of error rates with its major purpose being to increase the utility and informativeness of error rates in general.

2 Sequential Theory

Error rates are often given either as frequency data (how many errors occurred) or as a percent of errors (what proportion of responses were, in some way, erroneous), with those items producing lower error rates being preferred over those with higher error rates. However, even examining simple sequences made up of 50% error, CCII, IICC, ICIC, where “C” refers to correct performance trials and “I” to incorrect performance trials, though having equal percentages and frequencies of error, can lead to distinct behaviors in human end-users. For example, recent studies have indicated that not only do these patterns influence instrumental conditioning (Capaldi & Miller, 2004), where a participant must perform an action to cause the trial to be labeled as correct or incorrect, but can also influence variables dealing with Pavlovian or classical conditioning as well (Miller & Capaldi, 2006).

In the example that follows, extended training, such that the subject learns the pattern of the responses (correct/incorrect) is assumed. In this sense, we are describing expert performance. For example, given the series CCII, subjects learn that their correct responses are followed by incorrect, or that success is followed by failure. Given this pattern, it would be assumed that once an incorrect trial was encountered, the subject would then anticipate further failure, allowing their effort to lag. In other words, the memory stimulus of having a correct response would come to predict future incorrect responses, $S^{\text{Correct}} \longrightarrow$ Incorrect. In the case of an IICC series, the opposite would be assumed to occur, as, here, incorrect trials are followed by correct ones, persistence in the face of failure is followed by success, $S^{\text{Incorrect}} \longrightarrow$ Correct. Thus, in the IICC series, persistent and continuous effort in the face of failure would be trained, which would not be the case in the CCII series. The ICIC series would be assumed to create an intermediary condition, where persistence would be trained but not to the extent of the IICC series, where more failure was encountered before success. It is important to note that the IICC, CCII, and ICIC series each have the same percent correct, the same error rate percentage, and have an equal number of successful and unsuccessful trials. It is the pattern of those responses which influence

consequent persistence. Here, a concept which can be called I-length, or the number of incorrect trials followed by a correct trial may be useful. With sufficient training, a would be expected in expert performance, longer I-lengths should lead to greater persistence in the face of constant incorrect responses (failure). As an additional, albeit extreme condition, a series such as IIIIIC (I-length 6), $S^{6\ \text{Incorrect}} \longrightarrow$ Correct, with only a 16.7% success rate would be expected to produce more persistence than an I-length 0 series (CIIIIII) with the same percentage of success, an I-length 3 series (IIICIIIC) with a 25% success rate, or a CCCCCCCC (100% success) series, $S^{\text{Correct}} \longrightarrow$ Correct, where no experience with incorrect answers occurs and where success is only followed by additional successes. In the 100% success series, if switched to a situation where incorrect responses occurred, as the incorrect responses would be novel stimuli with no training, the participant would be expected to not persistent and cease responding rapidly. Thus, it is possible, and has been shown, that under the proper conditions, even series which produce the lowest success rates can also produce the highest sustained persistence. Interestingly, this indicates that participants sometimes persist least in conditions where they have enjoyed the greatest success in the past, while they may persist, and thus put forth the most effort, in cases where, historically, they have had the most failure. This is in contrast to some predictions of various optimization theories which suggest that participants would be the most persistent in tasks which, historically, have given them the greatest percentages of successes over time.

Additionally, sequential theory (Capaldi, 1994) also suggests that if the subject receives very little training, such that the pattern of responses (incorrect/correct) is not yet learned, as when a novice undertakes a task, then the patterns described reverse in their ordinal position. Thus, in this condition, for example, the IIIC pattern would be expected to produce less persistence than an ICIC one. The basic concept here is one of IC transitions or, how many times has the subject been exposed to incorrect responses followed by correct responses. Early in training, therefore, the group with the most IC responses is expected to persist the longest. Thus, being trained ICICIC would lead to more persistence than an IIICCC series, because in the former, three IC transition occurred and In the latter, only one was trained. Only after sufficient training do the I-lengths supplant the IC transitions in determining ordinal persistence in the face of extinction (all incorrect responses).

Data that support such a sequential view range from animal studies (Capaldi & Miller, 2004) to human studies (Miller, Terry, & Johnson, 2005) from simple alleyway running in rats to software usage in humans, and, anecdotally, in aviation maintenance training regimens. In general, it may be stated that such an analysis can inform training regimens across a wide field of areas. The major effect of taking into account such sequential variables would be to better understand why participants persist as some tasks more than others, even when error rates/frequencies are the same or similar, as well as ensuring that training takes into account the sequential effects of error. For example, in tasks where persisting in a task might damage equipment or cause other untoward consequences, the training itself can be built around sequential contingencies designed to minimize persistence. In the case where persistence in the face of repeated failures would be valuable, such sequential contingencies can also be programmed as part of the training regimen. In each case, it would be important to determine if the subjects of the training were considered to be novices or experts so

that the training would emphasize IC transitions for novices, moving to a focus on I-length series for experts.

As one example of a training contingency, a simple machine system might be used. If the machine was unreliable or required a great deal of effort or steps to accomplish then training should include the worker laboring until the task was finished without interruption. In this way, an undetermined I-length would be conditioned. A common mistake in training novices in this situation would be having the expert interrupt them and finish the work, thus making it so that the novice is not conditioned to experience success (IIIII… only and not III…C) and comes to rely on the expert to finalize their labors. Similarly, an expert who has been trained in long I-lengths and so persists in accomplishing a task might make mistakes with a new process where continual unsuccessful attempts may damage the machinery, etc., and so would need to be re-trained in the new CCCIIII… contingency where incorrect responses or failure is only followed by more failure, etc. While overall percentages of error rate might be useful to distinguish worker's quality, only by examining the sequence of their errors would other useful information predictive of their behaviors in the face of machine/system failures be gathered.

3 Conclusion

Thus, it is hoped that such an introduction to sequential theory, IC transitions, and I-length will be of benefit to those whose needs include anticipating human responding in the face of failure or continued incorrect responses. While error rate as whole is a valuable measure, by examining the particular sequences that make up the overall error patterns, error rate becomes at once more predictive and utilitarian.

References

1. Capaldi, E.J.: The sequential view: From rapidly fading stimulus traces to the organization of memory and the abstract concept of number. Psychonomic Bulletin & Review 1, 156–181 (1994)
2. Capaldi, E.J., Miller, R.M.: Serial learning in rats: A test of three hypotheses. Learning and Motivation 35, 71–81 (2004)
3. Miller, R.M., Capaldi, E.J.: An Analysis of Sequential Variables in Pavlovian Conditioning Employing Extended and Limited Acquisition Training. Learning and Motivation 37, 289–303 (2006)
4. Miller, R.M., Terry, R., Johnson, E.: An analysis of error rate as a predictor of user persistence in machine systems. In: Proceedings of the 11th International Conference on Human-Computer Interactions, vol. 4, Lawrence Erlbaum Press, Mahwah (2005)
5. Miller, R.M., Olds, K., Benson, J., Strong, N.: Question Order Effects on Task Persistence in Computerized Testing. In: Proceedings of the 10th Annual Global Chinese Conference on Computers in Education (2006)

Multidimensional Evaluation of Human Responses to the Workload

Shinji Miyake[1], Simpei Yamada[1], Takuro Shoji[1], Yasuhiko Takae[2], Nobuyuki Kuge[2], and Tomohiro Yamamura[2]

[1] School of Health Sciences,University of Occupational and Environmental Health, Japan
1-1 Iseigaoka, Yahatanishiku, Kitakyushu 808-8555, Japan
[2] Technology Developing Division, Nissan Motor Co., Ltd.
1-1, Morinosatoaoyama, Atsugi-shi, Kanagawa, 243-0123, Japan

Abstract. Changes in physiological parameters during mental tasks frequently show individual differences and discrepancies among responses invoked by mental tasks. A multidimensional analysis was used in this study to investigate these variations. Fifteen male participants performed a 5-minute multi-attribute task battery three times with different levels of difficulty. Principal component analysis was used for seven autonomic nervous system parameters recorded in five blocks including before and after resting periods. The first and the second principal components were plotted on the two-dimensional plane and their patterns were investigated. The results suggest that this method may provide more information on physiological responses induced by mental tasks.

Keywords: Workload, Autonomic Nervous System, Principal Component Analysis, MATB.

1 Introduction

Many research works in which several physiological parameters were measured to evaluate workloads have been reported. We also have investigated many physiological parameters to find sensitive indices to the workload [1], [2], [3], [4]. In such studies, we frequently encountered the problem of individual differences. For example, some participants showed bradycardia during mental tasks while others showed tachycardia [5], [6], [7]. Therefore, if we focus on only one specific index of physiological responses, e.g., heart rate, we may observe completely opposite responses induced by the same mental task. This study investigated variations of patterns of physiological responses. A multivariate statistical analysis (Principal Component Analysis) was used in a multidimensional examination of physiological responses obtained during mental workload experiments.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 379–387, 2007.

2 Method

2.1 Participants

Fifteen male undergraduates (age range 18-24 years; mean 21.1) participated in this experiment. All participants provided written informed consent and were paid a constant amount of money despite their task performance.

2.2 Procedure

The experimenter explained the detailed contents of the experiment and the tasks that the participants are required to perform. Then, written informed consent was obtained. All participants have done a practice session that contained three 5-minute trials of the medium level of difficulty before the experiment. Participants were given tips of this task to learn it easily.

On arrival in the laboratory on the experiment day, following physiological hookup and device check, each participant was left in a soundproof and electrically shielded room in a relaxed condition for 10-15 minutes to accustom the experimental situation. Then, he completed the following blocks: rest with eyes open before tasks (5 min: PRE), H (5-min: High level of difficulty), M (5-min: Medium level of difficulty), L (5-min: Low level of difficulty) and rest with eyes open after tasks (5 min: POST). This procedure was repeated three times on different day with at least one day interval. The order of task difficulty level was not randomized and set to the same order, i.e., H, M, and L [8]. Only the data on DAY1 were analized in this study.

2.3 Task

A revised version of the Multi-Attribute Task Battery (MATB) [9] was used. This task consists of tracking (first order, one dimensional compensatory tracking task, Figure 1, top center), system monitoring (Fig. 1, top left) and fuel management (Fig. 1, bottom center). The tracking task of the original version is a two-dimensional tracking task. It was revised to a one-dimensional (horizontal) tracking task in which a steering-type controller (CH Products, Virtual Pilot) was used. An amplitude ratio of forcing functions of the tracking task was set to 4:2:1 for high level (H), medium level (M) and low level (L) respectively.

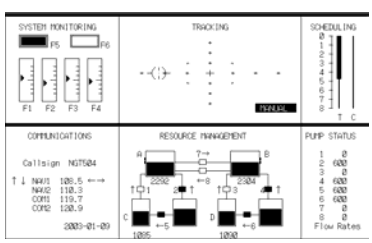


Fig. 1. NASA Multi-Attribute Task Battery display. It was negatively displayed in this experiment.

Participants were instructed to keep fuel levels of four tanks above specific fuel levels (2000 for Tanks A and B; 1000 for Tanks C and D) in the fuel management task. This instruction was different from the original one in which the participants were required to maintain the levels of tanks A and B at the optimum level of 2500 units each. Other tasks were almost identical as the original version. The descriptions of the three tasks were given in literatures [10], [11].

Harmonic means of reaction time for significant deviations in the four vertical gauges in the monitoring task were calculated as task performance indices. Also, average fuel levels of tanks A, B, C and D (sum of these four tanks) in the fuel management task and RMSE (root mean squared error) in the tracking task were recorded.

2.4 Physiological Measurements

Tissue blood volume (TBV) from the tip of the nose was measured using a Laser Doppler Blood Flow Meter (OMEGAWAVE OMEGA FLOW FLO-C1). Skin potential levels (SPL) from both palms were recorded by a DC amplifier (Nihon Koden, AG-641G). Average levels of these signals for each 5-minute block were calculated. The heart rate variability indices (LF: Low Frequency component, HF: High Frequency component, LF/HF, total power and CV-RR: coefficient of variation of RR intervals) and heart rates were calculated from the ECG signals recorded from CM_5 lead. The respiration signal using a strain gauge around the chest and hemodynamic parameters (SV: stroke volume, CO: cardiac output, PEP: pre-ejection period, LVET: left ventricular ejection time, and PL: preload) were also acquired (NEC Nicoview PA1100).

2.5 Subjective Measurements

The NASA Task Load Index [12] was used to evaluate some self-report assessments of task loading. This assessment technique contains six subscales (MD: mental demand, PD: physical demand, TD: temporal demand, OP: own performance, EF: effort and FR: Frustration). Furthermore, one item, OA (Overall workload), was added. The mean (RTLX: Raw TLX) and weighted mean (WWL: weighted workload) of these six scales were calculated.

Stress and arousal scores using the SACL (Stress Arousal Checklist) [13], state anxiety scores using STAI-S (State and Trait Anxiety Inventory) [14] and two mood scores (F: Fatigue, and V: Vigor) using POMS (Profiles of Mood States) [15] were obtained. The TLX paired comparisons and TLX ratings using visual analog scales were presented on a sub-computer screen at the end of each task trial. Rating scales for the mood evaluation were also appeared after resting period (before H trial) and after each TLX assessment.

2.6 Data Analysis

The following seven physiological indices that are not subordinate to each other and are supposed to be valid workload indices were selected and analyzed.

(1) Heart Rate (HR): HR itself may not be a sensitive workload index. However, a few participants showed pattern2 response in which HR decreased during mental tasks [7]. Therefore, HR was selected to investigate this response pattern.
(2) CV-RR: This is one index of heart rate variability and almost equivalent to the total power of an HRV power spectrum.
(3) Power spectral component of respiratory sinus arrhythmia (HF): Previous workload research reported that HF decreased due to the inhibition of vagal tone (parasympathetic nervous system) invoked by taskloads. Berntson et al. [17] who proposed two-dimensional autonomic space use this HF component (they describe this component as RSA) as the parasympathetic scale.
(4) PEP: This index was also used as the sympathetic scale by Berntson and his colleagues.
(5) SPL. (6) TBV. (7) SV. These three indices showed good correlations with workloads in our previous studies.

Above-mentioned seven parameters on DAY1 were standardized in each participant (five blocks: PRE, H, M, L, POST). Three principal components (first, second and third) were obtained using the Principal Component Analysis (PCA with varimax rotation). The first and the second components were investigated and average scores of twelve participants were calculated because PEP and SV from three participants out of fifteen were not acquired. The outline of this method was shown in Fig. 2. Results in subjective and performance data are not mentioned in this paper. They are reported in somewhere [7], [11].

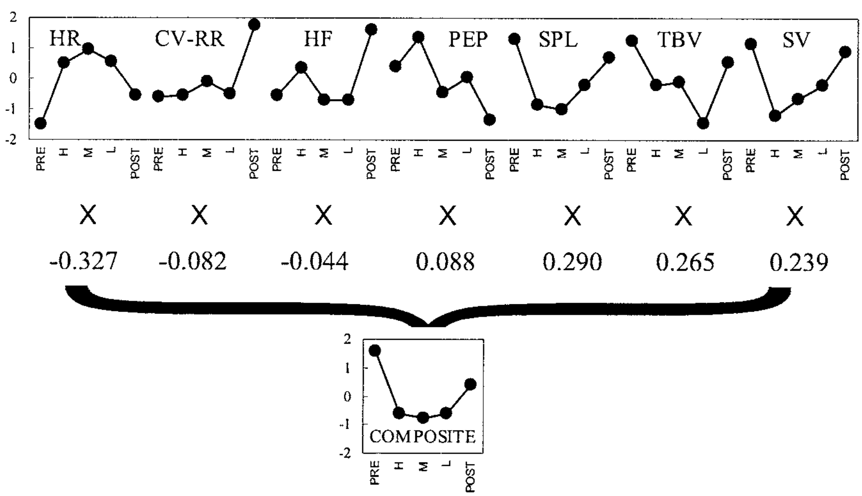


Fig. 2. Procedure to make a composite score (principal component). This figure shows an example of one participant (S1). Changes of seven physiological parameters (all of them were standardized) are shown at the top. Seven numerical values displayed in the middle are first principal component coefficients.

Repeated measures ANOVA and the post hoc multiple comparisons test (Student-Newman-Keuls) were applied and statistical significance was set at alpha = 0.05. The data were analyzed using SPSS for Windows, 11.0J.

3 Results and Discussion

The profiles of response patterns of seven physiological parameters are shown in Fig.3. The PCA was applied and three principal components were obtained to reduce these multivariate data. This procedure is similar to the analysis in the SD (Semantic Differential) method.

The average of the first principal components is shown in Fig.4a. Significantly higher scores are shown in before and after resting periods than task blocks. However, no significant difference was found among three task trials. The pattern of the second component of one participant (S3) is similar to the patterns of the first components of other participants. Therefore, we calculated the average scores of the first components using his second component. As shown in Fig.4b, the score in L-trial is significantly higher than other two task trials (H and M). This result suggests that selecting "proper" components when we calculate the average scores of principal components may make the difference clearer.

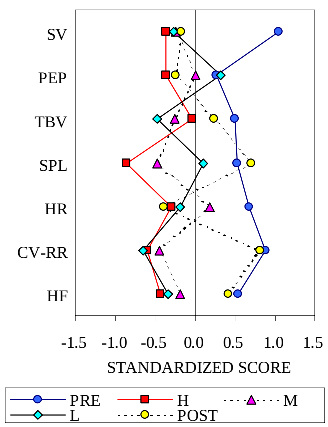


Fig. 3. Profile plots of seven physiological parameters. Scores are the averages of twelve participants.

The two-dimensional plots of the first principal component (horizontal axis) and the second principal component (vertical axis) of seven physiological parameters in each participant are shown in Fig.5. It is apparently shown that the second component scores in PRE resting period are different from POST resting period. However, the averages of the second component scores show no significant difference between PRE and POST resting period as shown in Fig. 6a. This is because the data of resting periods were plotted on the opposite quadrant of the two-dimensional plane in some

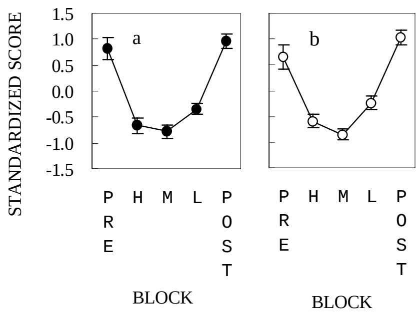


Fig. 4. Average of the first principal components (a: n=12). The second component was used instead of the first component for one participant (S3) to calculate the average score (b). The score in the L-trial in the right panel (b) is significantly higher than the H and M-trials although no significant difference was found among trials in the left panel (a). Bars represent standard errors of the mean.

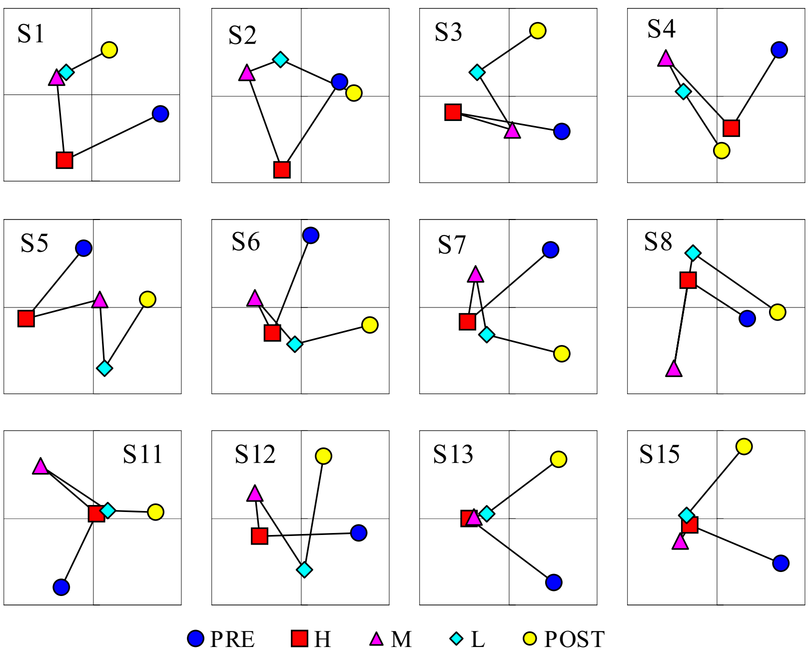


Fig. 5. Two-dimensional plots of the first components (horizontal axis) and the second components (vertical axis) of seven physiological parameters. Scales on horizontal and vertical axis are identical for all figures: -2 to +2. S1-S15 represent participants' codes.

participants (S4, S5, S6 and S7). When we adjust the vertical direction (signs of the second principal component scores) of them and calculate the average scores, the plot pattern clearly shows the large difference between PRE and POST resting periods

(Fig. 6b). The multiple comparisons revealed that PRE<M=H=L and PRE<POST=L ($p<0.05$. Fig.7). These results indicate that the physiological states in resting period after mental tasks are different from the resting period before the tasks. This is a very reasonable result but the first principal components could not detect this difference.

The bivariate (two-dimensional) autonomic space was proposed by Berntson, Cacioppo and Quigley [16]. PEP and RSA are allotted horizontal-sympathetic axis and vertical-parasympathetic axis in some research using this bivariate autonomic model [17], [18], [19]. Our data of PEP and HF (RSA) are plotted on this autonomic space in Fig. 8. No significant main effect of BLOCK was found in PEP. As clearly shown in Fig. 8, no significant difference was found between PRE and POST resting periods

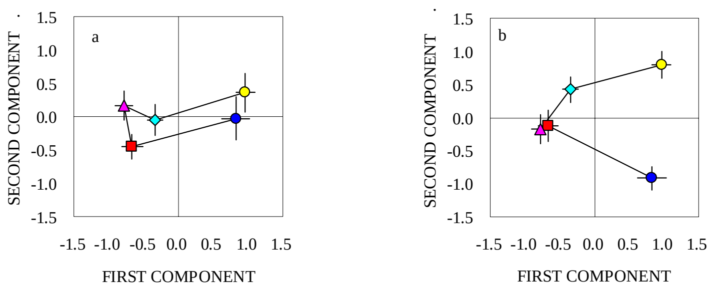


Fig. 6. Two-dimensional plots of the average scores of the first and the second principal components (n=12). No significant difference was found between PRE and POST in the second components in the left figure (a). When the signs of the second components of some participants (S4, S5, S6 and S7) were adjusted, the average scores show clear difference between blocks (b) as also shown in Fig.7. Bars represent standard errors of the mean. See the legend of Fig. 5 for marker patterns.

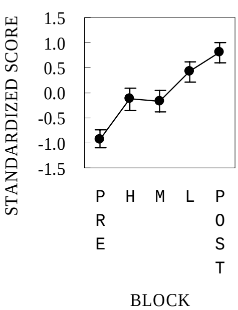


Fig. 7. Average second principal components (n=12). The signs of scores of four participants (S4, S5, S6 and S7) were reversed to adjust the response pattern. Bars represent standard errors of the mean.

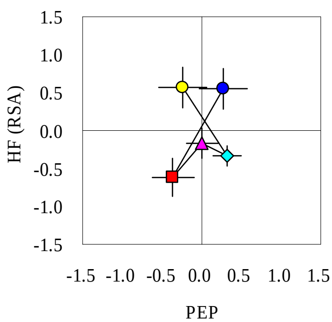


Fig. 8. Representation of responses on the bivariate autonomic space. Bars represent standard errors of the mean. See the legend of Fig. 5 for marker patterns.

whereas HF scores in both resting periods are significantly higher than the task blocks. Furthermore, no significant difference was revealed among task trials in HF. The results suggest that the bivariate autonomic space using PEP and HF is less sensitive to the workload than multivariate PCA in this study.

It must be clear that our purpose is not to evaluate autonomic nervous system (ANS) activities but to evaluate the mental workload. Using ANS parameters to evaluate mental workloads is very popular because some of those parameters are easy to measure in field studies. However, central nervous system (CNS) indices such as EEG and ERP (Event Related Potential) are also used to evaluate human responses to the workload. Recently, the NIRS (Near Infrared Spectroscopy) technology makes it possible to monitor brain activities (oxygen metabolisms) during task performances using the photo-sensors attached on the sculp surface. It goes without saying that our method to combine physiological parameters into one or two integrated indices by means of PCA is not limited to only ANS indices. We can combine not only ANS parameters, but also above-mentioned CNS indices, subjective workload scores, mood scores and performance parameters (note that performance data cannot be obtained during the resting periods) [20], [21]. However, the interpretation of vertical axes of figures 5, 6 and 7, i.e., the second principal components, is unknown at the moment although only ANS parameters are composed in this study. The second principal component coefficients for each physiological parameter may suggest the interpretation of the second component. Furthermore, subjective indices and mood scores are necessary to understand what this component means. Investigating the validity of this method considering above-mentioned more information is necessary to establish this method as the workload evaluation tool.

References

1. Miyake, S., Hashimoto, M., Miao, T., Shimizu, T., Tabe, M., Shiraishi, Y.: Effects of a Memory Task during a Multi-attribute Task. In: Adjunct Proceedings of 10th International Conference on Human - Computer Interaction (HCII 2003), pp. 73–74 (2003).
2. Miyake, S., Hashimoto, M., Miao, T., Shimizu, T., Tabe, M., Shiraishi, Y.: Effects of a Subsidiary Task on a Multi-attribute Task - Individual Differences in Responses. In: Proceedings of 10th World Congress and Exhibition on Intelligent Transport Systems and Services (CD-ROM), Paper #4084 (2003)
3. Miyake, S., Hashimoto, M., Shiraishi, Y., Takae, Y., Watanabe, T.: Human Responses to a Workload Change. In: Proceedings of Human Factors and Ergonomics Society 48th Annual Meeting 2004, pp. 2344–2348 (2004)
4. Miyake, S., Shoji, T., Takae, Y., Etori, N., Watanabe, T., Shiraishi, Y.: On-Road Investigation of Effects of Adaptive Cruise Control on Workload. In: Paper presented at Society for Psychophysiological Research 45th Annual Meeting, Lisbon, Portugal, September 21-24, 2005 (2005)
5. Sherwood, A., Turner, J.R.: A conceptual and methodological overview of cardiovascular reactivity research. In: Tutner, J.R., Sherwood, A., Light, K.C. (eds.) Individual Differences in Cardiovascular Response to Stress, pp. 3–32. Plenum Press, New York (1992)
6. Turner, J.R.: Cardiovascular Reactivity and Stress, Patterns of Physiological Response, pp. 51–54. Plenum Press, New York (1994)

7. Miyake, S., Yamada, S., Shoji, T., Takae, Y., Kuge, N., Yamamura, T.: Physiological Reactivity to a Multi-Task Environment. In: The 9th Korea-Japan Joint Symposium on Ergonomics, The Japanese Journal of Ergonomics, vol. 42 (suppl.), pp. 502–505 (2006)
8. Comstock, J.R., Arnegard, R.J.: Multi-attribute task battery. In: NASA-Tech. Memorandum 104174, NASA Langley Research Center, Hampton, VA (1992)
9. Molloy, R., Parasuraman, R.: Monitoring an automated system for a single failure: and task complexity effects. Human Factors 38(2), 311–322 (1996)
10. Nygren, T.E.: Framing of task performance strategies: Effects on performance in a multiattribute dynamic decision making environment. Human Factors 39(3), 425–437 (1997)
11. Miyake, S., Yamada, S., Shoji, T., Takae, Y., Kuge, N., Yamamura, T.: Physiological responses to workload change. A test/retest examination. In: Proceedings of 6th International Conference of Psychophysiology in Ergonomics (PIE2006), pp. 65–70 (2006).
12. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (TASK Load Index): Results of experimental and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 139–183. North-Holland, Amsterdam (1988)
13. Spielberger, C.D.: Anxiety as an emotional states. In: Spielberger, C.D. (ed.) Anxiety - Current trends and theory, Academic Press, New York (1972)
14. Mackay, C., Cox, T., Burrows, G., Lazzerini, T.: An inventory for the measurement of self-reported stress and arousal. Br. J. Soc. Clin. Psychol. 17, 283–284 (1978)
15. McNair, D.M., Lorr, M., Droppleman, L.F.: Profiles of Mood States. Educational and Industrial Testing Service, San Diego (1971)
16. Berntson, G.G., Cacioppo, J.T., Binkley, P.F., Uchino, B.N., Quigley, K.S., Fieldstone, A.: Autonomic cardiac control. III. Psychological stress and cardiac response in autonomic space as revealed by pharmacological blockades. Psychophysiology 31, 599–608 (1994)
17. Berntson, G.G., Cacioppo, J.T., Quigley, K.S.: Autonomic determinism: The model of autonomic control, the doctorine of autonomic space, and the law of autonomic constraint. Psychological Review 98(4), 459–487 (1991)
18. Berntson, G.G., Cacioppo, J.T., Fieldstone, A.: Illusions, arithmetic, and the bidirectional moduration of vagal control of the heart. Biological Psychology 44, 1–17 (1996)
19. Backs, R.W., Rohdy, J., Barnard, J.: Cardiac control during dual-task performance of visual or auditory monitoring with visual-manual tracking. Pshychologia 48(2), 68–83 (2005)
20. Miyake, S.: Multivariate KANSEI evaluation. In: Proceedings of the 2001 Spring Conference of Korean Society for Emotion & Sensibility, pp. 11–15 (2001)
21. Miyake, S.: Multivariate workload evaluation combining physiological and subjective measures. International Journal of Psychophysiology 40(3), 232–238 (2001)

The Influence of Visual Angle on the Performance of Static Images Scanning

Xiang Qiu[1,2], Yong Niu[3], and Xiaolan Fu[1,*]

[1] State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences,
Beijing 100101, China
[2] Graduate School of the Chinese Academy of Sciences,
Beijing 100049, China
[3] Beijing Jiao Tong University,
Beijing 100044, China
{qiux, fuxl}@psych.ac.cn, niuy2002@yahoo.com.cn

Abstract. The present study addressed to explore the influence of visual angle on the performance of static images scanning with a 2 (scanning distance) × 2 (scanning type) × 3 (visual angle) mixed design. The results demonstrated significant effects of three factors on participants' performance. Stimuli at 5.5° and 8.4° rather than 2.7° could facilitate the performance of static image scanning. However, the effect of visual angle on mental image scanning was smaller than on retinal image scanning. These findings were interpreted in terms of the theory of working memory and the theory of mental image. The implication of these findings in human-computer interface was discussed at last.

Keywords: visual angle, image scanning, mental load, working memory.

1 Introduction

Picture is one of primary external representations of the world as well as word [8], however, it supplies people with information in a quite different way from word. Compared with word, picture has the advantages of being an intuitionistic form, more understandable, and less susceptible to the influence of cross-cultural differences and literacy proficiency [15]. An abundance of pictures were used in human-computer interface [8], advertisement [15], traffic signs [1], and multimedia learning [10].

1.1 Retinal Image Scanning and Mental Image Scanning

When a picture appears in front of one's eye, a retinal image is formed through the light rays reflects from the picture activating the cells on retina, people then encode it as a mental image on visual working memory [7]. Both retinal image and metal image can subsequently be taken as an input of further processes such as scanning or rotating [13].

* Corresponding author.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 388–396, 2007.

Correspondingly, the image scanning can be divided into retinal image scanning and mental image scanning.

Retinal image scanning is the systematic shifting of attention over the retinal image of object or a scene with the presentation of the original real object or scene [7]. It is an integral part of visual search processes involving computer or radar displays, head-up displays in aircraft, or attending to road traffic signs [11]. Several variables were reported to affect the performance of users on retinal image scanning, including color [14], spatial frequency of picture [8], visual complexity [11], and number of icons and luminance of environment [15].

Mental image scanning corresponds to the systematic shifting of attention over the mental image when the real object or scene is out of sight [7]. Mental image scanning is an important spatial ability [2] and is often used in visual thinking [5]. An experimental measure of mental image scanning was developed by Kosslyn in 1973 and modified by Dror and Kosslyn in 1993 [2]. Details of this image scanning paradigm will be expatiated in later experiment section. By this modified paradigm, a lot of research found that more time was required to scan a greater distance across a mental image as well as across a retinal image.

A series of experiments had been reported by Kosslyn et al. reflected the structural isomorphism of visual mental images with corresponding retinal image [2] [7]. They also found both the retinal image and mental image share much of neural machinery [6]. However, most of them failed to give enough consideration to the difference between them. This is one aspect we focused on in the present study. According to Kosslyn's theory of mental image, both of retinal and mental image scanning work on the visual buffer, which is a functional space being located in the visuo-spatial scratch pad of working memory [7]. Subjects have to refresh their mental image on the visual buffer after the picture disappears, while retinal image need not such step because of presentation of the actual object. So, we hypothesized that retinal image scanning and mental image scanning might differ in the working memory resources they use. Scanning across a mental image demanded much more mental resources than retinal image scanning. We expected that there was some difference between the performance of the two types of image scanning, with higher accuracy and shorter response time for retinal image scanning than for mental image scanning.

1.2 Visual Angle and Image Scanning

Visual angle is defined as the angle between the line joining the observer to one point of the diameter of object and the line to the other point of the diameter of object. It is a function of object size and the distance between eyes and object (stadia distance). Visual angle reported in most researches was taken as a controlled variable while object size or stadia distance was taken as an independent variable (e.g. [2] [7]). However, it is visual angle but not object size or stadia distance that determines the size of retinal image. One obvious example is that the perceived size of object does not generally equal its physical size [5].

Though less data is available to reflect the relation between visual angle and image processing, a recent study carried out by Niu, Qiu, and Fu addressed the relationship between visual angle and maintenance of static images [12]. In their study, subjects firstly studied a 5×5 matrix that was composed of white and black panes, and then they were required to recall the locations of black panes. The results showed that the

matrix presentation with larger visual angle could promote the performance of short-term maintenance of static images. Given the fact that static image processing is based on the short-term maintenance of static images, we speculated in the present study that the performance of image scanning might also vary with visual angle.

To summarize, the current study was motivated by following questions. Firstly, was there a difference between the performance of retinal image scanning and mental image scanning? Secondly, whether visual angle could influence the performance of static image scanning? In order to answer these questions, Kosslyn's image scanning paradigm [2] and a variant of the paradigm were used to test the performance of mental image scanning and retinal image scanning respectively. Moreover, according to Huang et al's argument that acuity of vision was quite different among the three regions [0°, 5°], [5°, 8°], and [8°, 180°] [4], 2.7°, 5.5°, and 8.2° were taken respectively from these three regions.

2 Method

2.1 Participants

Thirty-six undergraduates participated in the experiment (18 men and 18 women), with a mean age of 22.5 years and an age range from 18 to 26 years old. All of them were right-handed and had normal or corrected-to-normal vision ability. All subjects were volunteers paid for their time.

2.2 Materials

As illustrated in Figure 1, a ring was constructed from twenty squares; with 6 squares on each side (two squares of them were blackened). A 0.5-cm-long arrow, corresponding to 0.6°of visual angle, appeared within the inside of the ring after the subject pressed the space bar (under mental image condition, this arrow was presented after the ring removed); the arrow pointed either north, east, south, or west. In addition, the arrow was positioned at two kinds of distances from the target square, either 0.4 cm away from it, or 0.8 cm away from it, which corresponded to 0.5° and 0.9°of visual angle, respectively. Totally 192 test trials were constructed, 96 for each

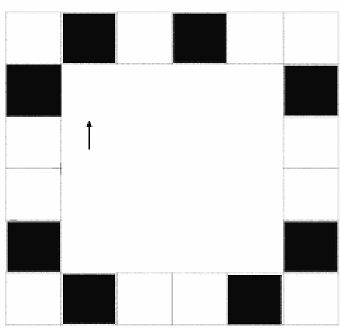


Fig. 1. Example of the stimuli used in retinal image scanning task

distance. On half of the trials at each distance, the arrow pointed to a black square, and on half it did not. The orientation of the arrow was counterbalanced within each condition. We also prepared 24 additional practice trials under each scanning condition.

2.3 Apparatus

The stimuli were displayed on a TCL17-in. monitor with a 70-Hz refresh rate at resolution of 1280 × 1024 pixels, connected to a Pentium IV computer. The experiment was programmed with E-Prime software (version 1.1). Participants positioned their heads on a chin located at a distance of 50cm from the center of the screen. Visual angle is calculated according to the formula as follows (α: visual angle; l: the diameter of the picture; d: stadia distance; $\pi \approx 3.14$):

$$a = 2\times \text{arctg}\,(l / 2\,d) \times 180\,\pi. \tag{1}$$

The sizes of the rings presented under the three visual angle conditions were 7.2 cm×7.2 cm, 4.8 cm×4.8 cm, and 2.4 cm×2.4 cm respectively with an unvarying stadia distance of 50 cm.

2.4 Procedure

First of all, the 36 participants were randomly allocated to each scanning condition, half of them for the retinal image scanning, and the other half for the mental image scanning. All the subjects were tested individually. Under the retinal image scanning condition, a trial began with a fixation point "+" at the center of the screen for 500 ms. Then the fixation was replaced by a blank screen for 250 ms. Following this, a square-ring stimulus with an arrow appeared at the center of the screen for 50 ms, and then the arrow disappeared leaving the square-ring on the screen. The subjects were required to scan from the pit of the arrow to the ring and pressed the yes key if the target square of the ring which the arrow pointed to was black or the no key if the target square was white. After the response, a cross appeared immediately, and a new trial began.

Under mental image condition, the subjects were shown a ring and were asked to study it until they could visualize it. Once they could visualize the ring, they pressed space bar and then the stimulus was replaced by a blank screen for 250 ms, following this, an arrow appeared for 50 ms. At this point, all the stimuli were removed. The subjects press the yes key if the arrow pointed to one of the black squares or the no key if it did not. After their response, a new trial would begin.

A 2 (scanning distance: 0.8 cm and 0.4 cm) × 2 (scanning type: retinal image scanning and mental image scanning) × 3 (visual angle: 2.7°; 5.5°, and 8.2°) mixed design was employed. The scanning type was a between-subjects factor, and the other two were within-subjects factors. All the trials were presented in random order; the same square-ring could not appear twice in consecutive trials. All subjects were asked to respond as quickly as they could while being as accurate as possible.

3 Results

Prior to analysis, response time 2 standard deviations away from the mean was omitted as outlier, which occupied 1.6% of all data. We did not include incorrect responses when calculating mean response time. Separate analyses of variance were performed on RT and ACC.

3.1 The Test of Speed-Accurate Trade-Off Effect

The analysis of Pearson's correlation between response time and accuracy rate was conducted to test whether there was a speed-accurate trade-off effect in the experiment. A significant negative correlation was obtained, $r = -0.615$, $p < 0.05$, suggesting less reaction time was associated with higher accuracy, so there was no significant speed-accurate trade-off effect in this experiment.

3.2 The Main Effect of Visual Angle and Scanning Type

The RT data were analyzed by repeated measure ANOVAs, which included scanning distance, scanning type, and visual angle of stimulus presentation as independent variables. In this analysis, as well as in every analysis reported in this article, all results not mentioned were not significant, $p > .10$.

As illustrated in figure 2a, subjects did require more time to scan a greater distance (with means of 859ms and 898 ms for near and far distances, respectively), $F(1, 34) = 5.49$, $p < .05$. The main effect of visual angle was found, $F(2, 34) = 21.315$, $p < .001$. Post hoc comparisons with LSD method revealed that this effect was due to large and medium visual angle conditions taking a significant less time to scan the same distance, $p < .01$. In addition, the difference between the two larger visual angles was not significant, $p > .10$. The main effect of scanning type was also in significance, less time was needed to complete the retinal image scanning task than mental image scanning task, $F(1, 34) = 25.00$, $p < .001$.

We also performed the corresponding analysis of the accuracy rates with the same independent variables. Consistent with the response time data, the subjects made more errors when a greater distance had to be scanned (with means of 86.1% and 73.8% for near and far arrow distances, respectively), $F(1, 34) = 163.07$, $p < .001$. The accuracy rates increased with increasing visual angle, $F(2, 34) = 5.80$, $p < .05$. Post hoc analysis indicated that only the difference of accuracy rates between the large visual angle condition and the small visual angle condition approached the significance, $p < .01$. The main effect of scanning type was also in significance, $F(1, 34) = 161.71$, $p < .001$. The mean accuracy rate of retinal image scanning was much higher than that of mental image scanning.

3.3 The Difference of Effect of Visual Angle Between the Two Types of Scanning

In the repeated measure ANOVAs, the most interesting findings were that the interaction between the visual angle and scanning type also approached significance,

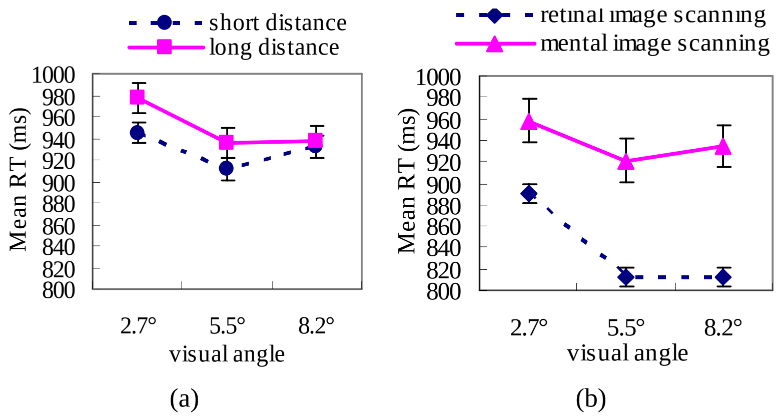


Fig. 2. Mean RT for each level of scanning distance (a) and scanning type (b)

$F(2, 68) = 5.43$, $p < .01$. A simple effect test showed the effect of visual angle was significant under both retinal image scanning condition and mental image scanning condition, $F(2, 68) = 28.74$, $p < .001$ for retinal image scanning and $F(2, 68) = 5.11$, $p = .009$ for mental image scanning.

In order to investigate the difference of effect of visual angle between the two types of image scanning, we calculated the mean decrease of RT when visual angle increased from 2.7° to 8.2° for each scanning type and compare the differences of mean decrease of RT between the two types of scanning. A one-way ANOVA revealed that the mean decrease of RT was larger under retinal image scanning (mean decrease = 77.28 ms) than under mental image scanning (mean decrease = 23.92 ms), $F(1, 34) = 10.34$, $p = .003$. When the visual angle increased from 2.7° to 5.5°, the mean decrease of RT was also larger under retinal image condition (mean decrease = 77.88 ms) than that under mental image condition (mean decrease = 37.28 ms), $F(1, 34) = 5.00$, $p < .05$. That is to say, there was a diminishing effect of visual angle from retinal image scanning to mental image scanning. The results were shown in Fig. 2b.

4 Discussions

The experiment reported here replicated the standard hallmarks of image scanning: scanning time and errors increased with increasing distance. We could have confidence that performance in these tasks did in fact reflect the efficacy of image scanning processes. In addition, we correlated the response time with accuracy rate of image scanning which revealed no speed-accurate trade-off effect in the experiment, so both response time and accuracy rate could be valid dependent variables. However, because of the sensitivity of RT, we took it as the main dependent variable as most studies on scanning task did [2] [7].

4.1 The Effect of Visual Angle

The results of the experiment demonstrated that the visual angle of stimulus presentation during the image scanning had a considerable effect on subsequent task performance. The two larger visual angle conditions facilitated the performance of both the retinal image scanning and mental image processing. These results were consistent with a previous study on the impact of visual angle on the short-term maintenance of static images [12], which found that larger visual angle could promote short-term maintenance of mental images. Huang, Cai, and Chen argued that the vision acuity would be maximal at the visual angle of about 5° according to the physiological structure of eye [4]. This was also in accordance with our results. In our experiment, best performance of image scanning tended to appear at 5.5° of the stimulus presentation, though there was no significant difference between 5.5° and 8.2°. However, since only three visual angles were used in our experiment, it was difficult to conclude that the larger the visual angle was the better image scanning task was performed. Future studies should be aimed to explore how visual angle influence the processing of different static images, and determine which visual angle will be the best to facilitate the performance of static image scanning.

4.2 The Difference Between Retinal Image Scanning and Mental Image Scanning

Clear differences between the performances of two types of scanning were also revealed by both the analyses of RT data and ACC data. Subjects under the retinal image scanning condition accomplished task with a much less response time and higher accuracy rate than those in mental image scanning. Previous studies paid more attention to the functional equivalence between the retinal image and mental image [2, 6, 7]. However, our experiment demonstrated that the retinal image and mental image were of some differences. Since we had excluded the contribution of extraneous variables by keeping the same color and shape of stimuli, the same orientation of the arrow and the same spatial frequency under the two conditions, it was easy for us to claim that the differences between retinal image scanning and mental image scanning was due to different mental load of the two kinds of scanning held. According to the theory of mental image raised by Kosslyn [7], all kinds of image scanning work on the visual buffer (which is a spatial mediate being located in the visuo-spatial scratch pad of working memory), so they are restricted by the visual working memory’s limited capacity. Under mental image scanning condition, subjects had to refresh their mental image on the visual buffer after the picture disappeared, while in retinal image scanning the picture didn’t disappear until subjects made a response, subjects only needed to maintain the orientation of the arrow. Hence, it was obvious that mental image scanning put a much greater mental load on subjects than retinal image scanning did.

Interestingly, the response time also revealed an unexpected interaction of scanning type and visual angle. Though simple effect test showed that the effect of visual angle existed on both the retinal image scanning and mental image scanning, we took a further analysis to test the difference of the decrease of response time from 2.7° to 5.5° or to 8.2° between retinal image and mental image scanning. The results indicated that the effect of visual angle was rather larger under the retinal image

condition than that under mental image condition. In other words, as the visual angle increased from 2.7° to 5.5° or 8.2°, a larger increase of performance appeared in retinal image scanning than in mental image scanning. One explanation for this interaction is the fact that retinal image, as a low level visual process, has a more direct relation with visual angle. Kosslyn's theory claims that mental image is a high-level visual process. Mental image scanning relies on such high-level mechanisms which typically operate on the output from low-level processes. It is possible that the influence of visual angle can be modified by a higher-level cognitive process. Further study should be taken to verify this hypothesis.

4.3 The Implications in Human-Computer Interface Design

These results may have important implications for human-computer interface design. For one thing, this study extends the visual factors which can influence the performance of image processing. In the realm of human-computer interface design, there are very few studies about the role of visual angle from psychology perspective. Though the effect of visual angle is similar to the effect of size [7], which refers to that a larger image could facilitate the processing of mental image, the effect of visual angle is of more practical applications and leaves a broader decision space for the designers of human-computer interface and computer & software manufacturers. Once a visual angle is fixed, the stimulus size can be altered flexibly with corresponding change in stadia distance. Secondly, icon search, one basic interaction between human and computer [8], is actually constructed by retinal image scanning. Many factors which can affect the performance of icon search have been reported, such as the number of icons and the contrast of icon [3]. Our research extends past studies by taking visual angle into consideration. The performance of icon search will be impaired if the visual angle of the stimulus presentation is too small. Lastly, not all the scanning tasks can be completed at the presentation of original stimulus, sometimes we have to complete some tasks by visualizing the original object. For example, designers often take three or four pictures to indicate successive procedures when they compile training procedures [10]. Only by mental image transformation such as mental image scanning can people make connection between one picture and the next one. So it is necessary to classify the scanning tasks into retinal image scanning tasks and mental image scanning tasks, and designers should pay more attention on the fact that some factors (like visual angle) may affect the two types of image scanning tasks in different ways.

Acknowledgments. This research was supported by grants from 973 Program of Chinese Ministry of Science and Technology (#2006CB303101), and the National Natural Science Foundation of China (#60433030).

References

1. Campagne, A., Pebayle, T., Muzet, A.: Oculomotor changes due to road events during prolonged monotonous simulated driving. Biological Psychology 68, 353–368 (2005)
2. Dror, I.E., Kosslyn, S.M., Waag, W.L.: Mental imagery and aging. Psychology and Aging 9, 90–102 (1993)

3. Fleetwood, M.D., Byrne, M.D.: Modeling icon search in ACT-R/PM. Cognitive System Research 3, 25–33 (2002)
4. Huang, X.T., Cai, Z., Chen, L.J.: The effect of visual angle on the recognition of Chinese character. Psychological Science 27, 770–773 (2004)
5. Foley, J.M., Ribeiro-Filho, N.P., Da Silva, J.A.: Visual perception of extent and the geometry of visual space. Vision Research 44, 147–156 (2004)
6. Kosslyn, S.M.: Mental Image and Human Brain. In: Jing, Q., et al. (ed.) Progress in Psychological Science around the World: XXVIII International Congress of Psychological Proceedings, pp. 185–195. East China Normal University Press, Shanghai (2006)
7. Kosslyn, S.M.: Image and Brain: the Resolution of Imagery Debate. MIT Press, Cambridge (1994)
8. Li, L.S.: Human-computer Interface Design, pp. 225–231. Science Press, Beijing (2004)
9. Mast, F.W., Ganis, G., Christie, S., Kosslyn, S.M.: Four types of visual mental imagery processing in upright and tilted observers. Cognitive Brain Research 17, 238–247 (2003)
10. Mayer, E.R.: Multimedia Learning. Combridge University Press, Combridge (2001)
11. McDougall, S., Tyrer, V.: Searching for signs, symbols, and icons: Effects of time of day, visual complexity, and grouping. Journal of Experimental Psychology: Applied 12, 118–128 (2006)
12. Niu, Y., Qiu, X., Fu, X.L.: The influence of visual angle and quantity of information on the short-term maintenance of static images. Ergonomics (in press)
13. Paivio, A.: Imagery and Verbal Process. New York: Holt, Reinhart & Winston (1971)
14. Vierck, E., Miller, J.: Direct selection by color for visual encoding. Perception and Psychophysics 67, 483–495 (2005)
15. Zhu, Z.X. (ed.): Engineering Psychology. People’s Education Press, Beijing (2003)

Occurrence of Secondary Tasks and Quality of Lane Changes

Laurence Rognin[1], Sophie Alidra[2], Clément Val[3], and Antoine Lescaut[4]

[1] LAB, Renault PSA Peugeot Citroën, 132 rue des Suisses,
92 000 Nanterre, France
[2] INTA, Boulogne Billancourt, France
[3] CEESAR, 132 rue des Suisses,
92 000 Nanterre, France
[4] PSA Peugeot Citroën, Vélizy Villacoublay
laurence.rognin@lab-france.com, sophie.alidra@inta.fr,
clément.c.val@ceesar.asso.fr, antoine.lescaut@mpsa.com

Abstract. Methods to assess driving degradation due to driver distraction are currently discussed and defined by international standardization groups. A simulator experiment involving 17 participants was conducted to assess the reliability and relevance of one of these methods (Lane Change Test) to discriminate between secondary tasks. In addition to varying age groups, vehicles and secondary tasks, the protocol was also varied to assess the impact of the instruction occurrence and its possible conflict with primary task performance. Results show the limitations of the main parameter proposed by the method (lateral deviation) and question the reliability of the method in its current form. Additional indicators seem necessary to make sense of the respective impact of the varied conditions. Surprisingly, the impact of the instruction occurrence is very limited, apparently because individual strategies have more impact than situations differences.

Keywords: Driver distraction, driving performance, Lane Change Test, secondary tasks, simulator experiment.

1 Introduction

With the continuous development of in-vehicles comfort, information and assistance systems, car manufacturers are facing series of antagonistic demands. First, the consumer demand for new technologies and innovations is counterbalanced with regulation authorities objectives to ensure a robust integration and a cautious usage of these systems. Second, the new applications aiming to reduce driver load (e.g. navigation systems) actually increase the risks of errors in introducing new sources of distraction.

The usage of in-vehicle systems while driving increases the multitasks demand and raises the question of efficient resources management enabling overall safety to be maintained. Typically, navigation systems designed to guide the driver between destinations require extensive and complex set-up, not to mention real-time follow-up

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 397–406, 2007.

of visual and/or verbal instructions. Even though these attractive systems provide benefits (comfort, social image, reduced risks of distraction due to map reading while driving), their impact on driver loss of attention and on safety needs to be assessed.

Involving the combination of multiple manual and cognitive resources, driving requires the driver to allocate appropriately and efficiently his/her limited capacities [1]. In addition to the motor skills required to maneuver and control a vehicle, cognitive resources are essential to perform driving-related tasks such as information perception, event detection, problem solving and decision making. Continuous monitoring of the surroundings (including through rear and side view mirrors), essential to maintain situation awareness requires the driver to focus attention to the road scene. Any glance inside the vehicle, i.e. away from the road scene, reduces the driver opportunity to detect hazards and handle them safely. Wickens [2] proposed a multiple resources management model to describe how driver could simultaneously handle this variety of tasks. This model highlights the driver ability to use concurrently different modalities (visual, auditory, tactile, …) to apprehend situation demand and perform efficiently and safely the driving tasks. Based on this theory, the impact of display clutter, location, modality on driver distraction has been measured using various metrics: scanning patterns, hazard detection time, quality of lane keeping, secondary task quality ([1], [3], [4], [5]). Some computational models are also currently investigated to predict the amount of interference that will occur between two or more tasks performed concurrently.

The recent naturalistic study conducted in US suggested that driver's glances away from the forward roadway potentially contribute to a much greater percentage of events than has been previously thought, with 78% of the crashes and 65% of the near crashes [6]. Secondary-task distraction contributed to over 22 percent of all crashes and near-crashes [7], which is in-line with recent French results [8]. Whereas research studies are essential to provide a better understanding and knowledge of the driver (e.g. strategies, capabilities and limitations), car manufacturers face a pressing need for fast, simple, cost-effective and reliable method to measure the potential impact of new in-vehicle systems on driver distraction and safety. Risks induced by driver distraction vary with the type, timing, intensity, frequency and duration of this distraction. It is essential to understand these factors and their mutual influences. Several methods currently discussed at an international level (ISO TC22/SC13/WG8) are intended as "tools to help system designers ensure that the intended benefits outweigh the risks of devices and features that are meant to be used while driving" [9]. In this context, a method denoted Lane Change Test (LCT) aims at measuring quantitatively the degradation of driving performance induced by secondary tasks. Previous experiments conducted in the LAB proposed improvements in terms of experimental protocol (e.g. vehicle-based protocol) and analysis (e.g. individual reference trajectory, eye-tracker data, position on lane).

To build on efforts to assess the LCT method ([10], [11]) a new experiment was conducted on a simulator in autumn 2006. The main objectives were to assess the relevance and robustness of the LCT method, to identify its main limitations and if necessary refine it.

2 Method

2.1 Participants

Seventeen participants of two age groups ([25-54] and [60-70]) were recruited through public notice. All had valid driver's licences, a minimum of 4 years of driving (mean=28 and max=48) and drive on average 16000 kilometers per year (min=5000 and max=25000). The same participants were involved in the four successive sessions.

2.2 Apparatus

Equipment. Four different production vehicles were tested. Attention was paid to ensure that the systems tested were comparable in terms of functions provided and modalities of interaction (input and display). The vehicles were positioned in front of a 2x3 meters video screen where the driving scene was projected. Front wheels of the test vehicle were placed on swiveling plates to reduce friction to ground. Video camera were placed inside the vehicle to collect three complementary views: driver's face, over the shoulder view and HMI view. Additional markers were provided to enable the experimenter to highlight events of interest (e.g. beginning / end of secondary tasks). Scenario and recording (system and video) were automatically launched from the experimenter workplace. For a calibration task, a dedicated 15" monitor was positioned on the right side of the route scene and a simplified keyboard (limited to arrow keys) was used to perform the designation and selection task. When not necessary this display was removed from the scene. For the other secondary tasks (radio manipulation, interaction with the navigation system), displays available in the tested vehicles were used.

LCT Software and task. The tool developed in the context of the ADAM project [12] was used to perform the Lane Change Test. The Lane Change Test (LCT) is a simple laboratory dynamic dual-task method that aims to quantitatively measure performance degradation on a primary driving-like task while a secondary task is being performed. The LCT comprises a simple driving simulation that requires a test participant to drive along a straight 3-lane road at a constant, system controlled, speed of 60km/h. Participants are instructed in which of the lanes to drive by signs that appear at regular intervals on both sides of the road. Participants use the vehicle steering wheel to maintain the position of the simulator vehicle in the centre of the indicated lane and are prompted to change lanes according to the instructions on the signs. The only visual feedback the participants get is the front view (i.e. no rear nor side view provided in mirrors). Engine sound was simulated to increase situation realism. The scene consisted of a series of 3 km test tracks, with lane change signs displayed every 150m. Participants had to perform maneuvers as quickly and efficiently as possible. Actions on the steering wheels were instrumented and transmitted to the simulation tool in order to reproduce on screen lateral changes.

2.3 Experimental Design

Run plan. For each vehicle tested, the experiment used a 2 (age group: medium, senior) x 5 (secondary task: none, calibration, radio scrolling, radio list and navigation) x 3 (occurrence: at the sign, 50m before, 50m after) repeated measures design.

Secondary tasks. To enable comparison between LCT studies, the Surrogate Reference Task (SuRT) was used as a calibration task (standardized reference). It required the participants to locate a target among visually similar distractors (visual demand) and then select the portion of screen containing the target (manual demand). In addition, three other tasks were tested in each of the four vehicles: radio frequency scrolling, radio station selection and destination entry in the navigation system. The radio scroll and the radio list were very similar in all vehicles. The navigation tasks differed both in terms of navigation in menus and accessibility of interaction devices: input devices were located on the front panel for vehicles 1, 2 and 4 and on the right side of the driver for vehicle 3. To avoid boredom, radio and navigation tasks were mixed and occurred between 1 and 2 times each within each track. To ensure comparable conditions between subjects and between successive vehicles, secondary tasks instructions were pre-recorded and automatically issued at a same moment defined in distance to lane change sign (-50m, 0m or +50m). The three occurrence correspond to the appearance of the lane change sign, the end of the lane change and the maintaining of the trajectory.

Programme. Prior to the experimentation, all participants tested the experimental set-up, essentially to ensure that none of them suffered from the simulator sickness. Four different sessions of two weeks each were organized between September and November 2006. For each vehicle, every participant went through sessions of two hours, including training, measures and debriefing. Each of the four sessions began with a training period, whose objective was for the participants to become familiar with both the primary (drive and change lanes) and the secondary tasks. For the measured runs, the participants drove along 10 successive tracks: without secondary task (tracks 1 and 10), with calibration task (tracks 2 and 9) and with mixed secondary tasks (tracks 3 to 8).

Expected LCT output. The method was expected to be sensitive to *secondary tasks* (less degradation with radio than with navigation tasks) and *instruction occurrence* (more degradation before the sign due to disturbance in sign detection). No difference between *vehicles* was assumed, but the method should expected to be too sensitive to *drivers experience* with the simulation tool and consequently not reliable enough to compare successively different vehicles with same participants.

3 Results

The objective and subjective data collected consisted of vehicles parameters, LCT simulator logs, experimenter's markings, audio and video recording of participants'

actions and comments, experimenter's observations, interviews and questionnaire items. In the present paper, the focus is on medium age group results.

3.1 Lane Change Task Performance

To reflect the degradation of the lateral control possibly induced by the secondary task, the actual course is compared with a theoretical one. The mean lateral deviation "is the total area between the normative model and driving course (in m2) divided by the distance driven (m)." [9] The mean deviation in lane change was calculated per task in a repeated measures analysis. Typically, measures taken between beginning and end of a task were then summed to obtain a mean per type of task. Comparisons between means were made using 95% confidence intervals. Despite its interest, this normative model seemed too ideal, and often far from actual performances. A more accurate deviation was calculated on the basis of participants average lane changes (initiation of the change, rate of change) in the baseline condition. Similar trends were observed with normative and adapted deviations, but deviation values were smaller when using the adapted model (Fig. 1 left). Participants showed a greater mean deviation when performing two tasks (e.g. driving and manipulating radio) than when performing the baseline condition (driving without secondary task) [$F(4,414)=57.54$, $p<0.001$]. Performances were similar in the drive only and calibration conditions. Mean values of the adapted deviation were similar in all conditions. Surprisingly, the task estimated as the most difficult (navigation) induced less deviation than the two other tasks (radio list and radio scroll). The participants showed a larger normative deviation when scrolling frequencies, but no difference between tasks in terms of adapted deviation. The mean deviation metric (normative and adapted) does not seem sensitive enough to discriminate between tasks within the medium group. No significant difference between the vehicles is observed.

The impact of the instruction occurrence varies along with the type of secondary task: the worst result (larger deviation) is obtained with the "0" occurrence for the

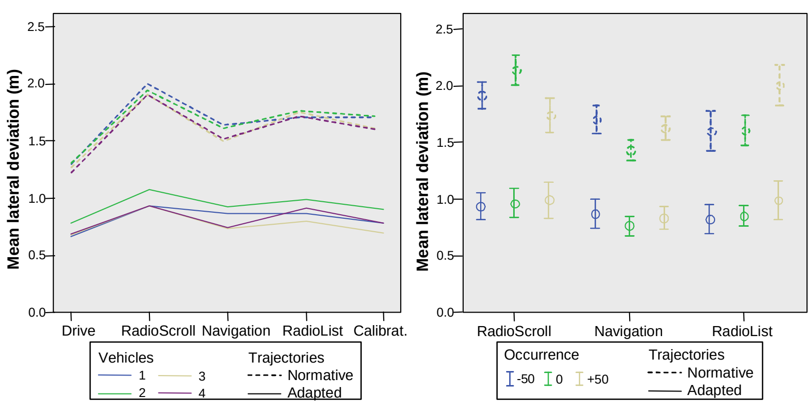


Fig. 1. Lateral deviation as a function of vehicle (left) and instruction occurrence (right) for the medium age group

radio scroll and "+50" for the radio list. The best result is obtained with the "-50" for the navigation (Fig. 1 right). The results suggest that each secondary task affects differently the various dimensions of the primary task (e.g. sign detection, change initiation, change maneuver, position adjustment), each corresponding to a specific position on the trajectory.

3.2 Secondary Task Performance

Even though the standardized LCT is limited to the analysis of the deviation metric, it was decided to consider additional indicators and assess their potential added value. The secondary tasks were characterized in terms of their duration and latency and compared according to the age, vehicle and occurrence factors. To calculate duration and latency, the start of action was defined as the first action on the device. The navigation task was the longest (between 50 and 60 seconds), while radio tasks were much shorter (20-30 seconds for the radio scroll and 15-20 seconds for the radio list). Participants showed the largest latency for the radio scroll task, possibly due to the structure of the instruction, with the relevant information at the end of the message (Fig. 2). The latency values were larger with the first vehicle. This suggests a learning effect: the participants gradually learnt to anticipate the tasks, initiating actions even before the end of the instructions. Both duration and latency of the navigation task were increased with the "- 50" condition (Fig. 3). Two explanations are proposed: (i) one can assume that in this condition, participants waited for the sign to appear before initiating this long task; (ii) this could reflect the strategy used to perform the task, corresponding to interruptions to focus momentarily on the lane change task. This last point could also explain why the most efficient lane change task performance (reduced deviation) is observed with this navigation task. No other effect of the occurrence on the secondary tasks performance was observed. A possible explanation lies in the differences in individual strategies. Two main categories were identified: "in a hurry", who initiates the task as soon as the instruction is issued and perform it quickly and "careful" who gives priority to driving, wait for the appropriate not conflicting moment to initiate it, perform it cautiously and occasionally interrupt it. Each instruction occurrence might finally correspond to a different situation on the track for participants according to the strategy they use: "-50" for a "careful" participant might correspond to "0" for a participant "in a hurry".

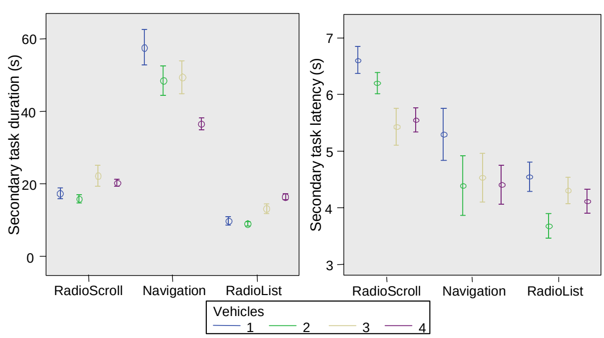


Fig. 2. Duration and latency of secondary tasks, medium age group, all vehicles

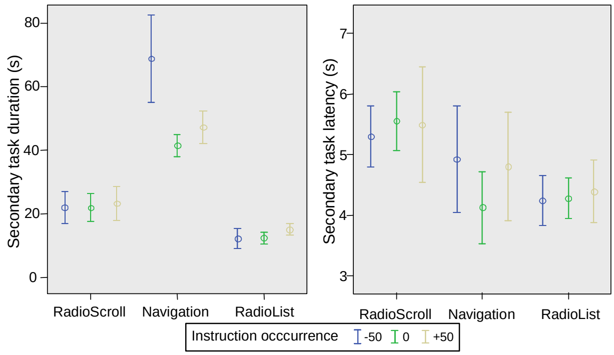


Fig. 3. Duration and latency of secondary tasks, as a function of instruction occurrence, medium age group, vehicle 3

4 Discussion

4.1 Inter-individual Differences

Neither normative nor adapted deviation was sensitive enough to discriminate between tasks and both provided surprising results, with the longest and most complex task measured as the less distractive in terms of driving degradation. This questions the SAE Recommended Practice J2364, commonly known as the 15-Second Rule for Total Task Time or the 15-Second Rule, which specifies the maximum time allowed (15 seconds) for completing a navigation system task involving manual controls and visual displays when the task is performed statically [13]. Latency and duration of secondary tasks provided discrimination between tasks and driver strategies. Observations of the participants practices, combined with a more detailed analysis of the instructions used revealed a major difference. For the radio scroll, the participants had to wait until the end of the message (5 seconds later) to know the frequency and decide on which arrow (up or down) to press, whereas they could initiate actions earlier for the radio list (press on "mode" or "list" button) and for the navigation tasks. The latency in secondary tasks initiation results in a different position on the trajectory, and consequently conflict between different resources allocations. The various strategies observed during the experimentations were consistent with previous findings on cognitive heuristics in multitasking environment [14]. Because of the varied strategies, the moments of occurrence needs to be more precisely characterized for each driver profile in clarifying the expected conflicts (e.g. sign detection, position adjustment, trajectory maintaining). Typically, the latency of radio scroll task results in 10 meters "delay", which may correspond, with the second occurrence ("0") and a "hurry" profile, to the most demanding task in terms of visual and manual control. In terms of visual load, the driver needs to look both at the radio device to read the frequencies and at the road to appropriately end the lane change and adjust his/her position on the new lane. In terms of manual load, the driver must both manipulate the radio (press on arrows) and move the steering wheel to adjust position. The duration of tasks also influences strategies: on the one hand, the navigation task,

considered by participants as long and complex was usually initiated quickly by the drivers but decomposed in successive independent sub-tasks. This resulted in conflicts with the detection of the change sign in the “-50” occurrence. On the other hand, the radio list, considered as short and simple resulted in lower pressure in the “+50” condition (no sign to detect, no adjustment on lane) and led drivers to literally dive into the task, focusing their whole attention on the radio device. To go a step further, drivers’ eye movements in similar situations should be analysed, as this raises the issue of situation awareness and driver capabilities to detect events and react appropriately.

4.2 Methodological Issues

The observations during experiments, coupled with the analysis of actual trajectories showed compensation actions at the end of secondary tasks. Typically, after the last action (i.e. after the “end” marker), the driver adjusts his/her course to replace the vehicle in the middle of the lane. In the current analysis, those periods are not covered. This raises the question of the definition of beginning and end of secondary tasks and associated periods of analysis. The analysis of secondary tasks latency and duration also raises the issue of the learning process. The gradual reduction of latency suggests that with practice participants get familiar with what is expected and confident with their ability to initiate tasks. Typically, they learn with practice that for navigation and radio list tasks they can initiate actions even before the end of the verbal instructions. This effect questions the involvement of same participants in series of studies investigating successively different systems. Combined with the observation of different driver strategic profiles (quick versus careful), this raises the issue of participant selection and experiments reproducibility.

Even though the LCT method was never designed to reproduce the reality of driving (but rather to enable comparison between systems), it can be emphasized that the simplified setting limits the driving to the lateral control of the vehicle and the detection of information located in front of the vehicle. The allocation of visual resources observed in the LCT settings and the associated measures of driving degradation need to be put into perspective. Indeed, the impact of the distribution of visual attention between two sources located in front view might be completely inadequate when this attention needs to be split between opposite directions (front versus side or rear). As we could see, even though the LCT indicator (lateral deviation) is not sensitive enough to discriminate among medium age participants, it provides interesting results (longest tasks inducing lesser deviation). Typically, it suggests that the interruptibility of the task is not inherent to the task, but rather reflects a user intention. To improve the analysis and interpretation of results, the method could benefit from additional perspectives. Typically, expert evaluation or user testing of the in-vehicles systems used for secondary tasks could provide relevant information in terms of secondary tasks characterization, strategies and trade-offs implemented in the course of action. The tasks simulated with the LCT tool are realistic in the sense that guidance, lateral control, lane change are real driving tasks. However, the frequency of the tasks is not realistic. The question of generalization of results needs to be questioned. Again, one needs to remember the initial objectives of the method and consider carefully strategies and performances obtained in these

conditions. Whereas an improved indicator might enable different design options to be compared it should not be envisaged as a means to evaluate one option per se.

5 Conclusion

The simulator experiment conducted with two age groups (senior and medium) assessed the relevance and the reliability of the LCT method to measure degradation of the driving task due to secondary tasks performance. In addition to varying age groups, vehicles and secondary tasks, the protocol was also varied to assess the impact of the instruction occurrence and its possible conflict with primary task performance. Results show the limitations of the main parameter proposed by the method (lateral deviation) and question the reliability of the method in its current form. Additional indicators seem necessary to make sense of the respective impact of the varied conditions. Surprisingly, the impact of the instruction occurrence is very limited, apparently because inter-individual differences have more effects than the situations differences. The participants seem to implement different strategies according to the specificity of tasks: short tasks induce a “dive in” behavior, with an intense but brief focus on the in-vehicle devices, whereas with longest tasks the driver attention is split between the driving and the secondary task. Beyond understanding what type of deviation is the most acceptable in terms of safety (i.e. between a large but short deviation and a small but long one), the next step will consist in investigating the quality of lane changes performed. To discriminate low quality of lateral control from errors in lane changes (omission or incorrect change), the type and frequency of errors are currently analysed.

Acknowledgments. The authors wish to thank the C.E.E.S.A.R technical team who prepared the experimental set-up and the drivers who took part in the experiment. Special thanks are also addressed to reviewers for their fruitful comments.

References

1. Wickens, C.D., Seppelt, B.: Interference with driving or In-vehicle task information: the effects of auditory versus visual delivery (Tech. Report AHFD-02-18/GM-02-3). University of Illinois, Aviation Research Laboratory, Savoy, IL (2002)
2. Wickens, C.D.: Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science 3(2), 159–177 (2002)
3. Horrey, W.J., Wickens, C.D.: Multiple resource modeling of task interference in vehicle control, hazard awareness and in-vehicle task performance. In: Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, pp. 7–12. Park City, Utah (2003)
4. Horrey, W.J., Wickens, C.D.: Driving and side task performance: The effects of display clutter, separation, and modality. Human Factors 46(4), 611–624 (2004)
5. Horrey, W.J., Wickens, C.D., Consalus, K.P.: The Distracted Driver: Modeling the Impact of Information Bandwidth, In-Vehicle Task Priority, and Spatial-Separation on Driver Performance and Attention Allocation (Technical Report AHFD-05-11/GM-05-2). University of Illinois, Aviation Human Factors Division, Savoy, IL (2005)

6. Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sudweeks, J., Perez, M.A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z.R., Jermeland, J.: (In press). The 100-car naturalistic driving study, Phase II – results of the 100-car field experiment (Contract No.DTNH22-00-C-07007). Washington, DC: National Highway Traffic Safety Administration.
7. Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., Ramsey, D.J.: The Impact of Driver Inattention On Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. In: Report DOT HS 810 594, National Highway Traffic Safety Administration, Washington, DC (2006)
8. Thomas, C., Le Coz, J.-Y., Page, Y., Damville, A., Kassaagi, M.: Car Driver inactivations in real-world precrash phase. In: Conference Convergence, Detroit, USA (October 2000)
9. ISO Working Draft (2006). Road vehicles – Ergonomic aspects of transport information and control systems — Simulated lane change test to assess in-vehicle secondary task demand. ISO TC22/SC13 N WG8 N533 (2006)
10. Burns, P.C., Trbovich, P.L., McCurdie, T., Harbluk, J.L.: Measuring distraction: task duration and the lane-change test (LCT). In: Proceedings of the Human Factors and Ergonomics Society 49th annual meeting, pp. 1980–1983 (2005)
11. Lindgren, A.: Navigating navigation. A safety and usability evaluation of the Volvo P1 navigation system. Master Thesis in Cognitive Science. Report LIU-KOGVET-D-05/22-SE, Linkoping University, Sweden (2005)
12. Mattes, S.: The Lane Change Test as a tool for driver distractionevaluation. IHRA-ITS Workshop on Driving Simulator Scenarios, October 2003 – Dearborn, Michigan (2003)
13. Society of Automotive Engineers (2000). Navigation and Route Guidance Function Accessibility While Driving (SAE Recommended Practice J2364), draft of January 20, 2000, Warrendale, PA: Society of Automotive Engineers (2000)
14. Kiefer, J., Schulz, M., Schulze-Kissing, D., Urbas, L.: Cognitive heuristics in multitasking performance. European Cognitive Science Conference (EuroCogSci), 2007 Delphi (submitted) (2007)

What Really Is Going on? Review, Critique and Extension of Situation Awareness Theory

Paul M. Salmon*, Neville A. Stanton, Daniel P. Jenkins, Guy H. Walker, Mark S. Young, and Amardeep Aujla

Defence Technology Centre for Human Factors Integration (DTC-HFI)
Brunel University, BIT Lab, School of Engineering and Design, Uxbridge, Middlesex, UK, UB8 3PH
paul.salmon@brunel.ac.uk

Abstract. Theoretically, Situation Awareness (SA) remains predominantly an individual construct. The majority of the models presented in the literature focus on SA from an individual perspective and in comparison, the concept of team SA has received less attention. SA in complex, collaborative environments thus remains a challenge for the human factors community, both in terms of the development of theoretical perspectives and of valid measures, and also in the development of guidelines for system, training and procedure design. This article presents a review and critique of what is currently known about SA and team SA, including a comparison of the most prominent individual and team models presented in the literature. In conclusion, we argue that recently proposed systems level Distributed Situation Awareness (DSA) approaches are the most suited to describing and assessing SA in real world collaborative environments.

Keywords: Situation Awareness, Teams, Collaborative Systems.

1 Introduction

Ostensibly, Situation Awareness (SA) refers to the level of awareness and understanding that an individual has of a situation, an operator's dynamic understanding of 'what is going on' (Endsley, 1995a). Such is its popularity, much has been written about the construct, yet there remains some disparity between the various models presented in the literature. Further, SA in complex collaborative environments requires much further investigation, despite the fact that team SA is now recognised as a critical factor in collaborative system design and assessment (Shu and Furuta, 2005). This article presents a summary of a review and critique of what is currently known on SA and team SA, following which a recently developed approach for describing SA in complex collaborative environments is presented.

2 Situation Awareness

First emerging as a concept of great interest in during the First World War (Press, 1986; cited in Endsley, 1995a), SA has since enjoyed great attention from the human

* Corresponding author.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 407–416, 2007.

factors community. Following Endsley's (1995a, b) seminal work, published in a special issue of the *Human Factors* journal on SA in 1995, SA rapidly emerged as a prominent factor in system design, analysis and performance-based research across many complex domains. Interest in the construct continues to expand, yet it remains a contentious topic. Depending upon where you look in the literature, there are various different views on the construct and there have many attempts at defining it. The main point of contention lies in the description of SA as either the cognitive *process* used to develop and maintain SA, to the tangible *product* of SA, or to a combination of both the process used and the end product of SA. Endsley (1995a), for example, describes SA as a *product* (resulting from a *process* called *situation assessment*) comprising the following three levels:

> *"The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future"* (Endsley, 1995a, p 36).

Fracker (1991), on the other hand, describes the *process* of "combining of new information with existing knowledge in working memory and the development of a composite picture of the situation along with projections of future status and subsequent decisions as to appropriate courses of action to take" (Fracker, 1991). From the *process-product* school, Smith and Hancock (1995) describe the construct as 'externally directed consciousness' and suggest that SA is, "the invariant in the agent-environment system that generates the momentary knowledge and behaviour required to attain the goals specified by an arbiter of performance in the environment" (Smith and Hancock, 1995).

3 Situation Awareness Theory

Inaugural SA theories were, in the main, focused on describing the SA of individual operators (e.g. Endsley, 1995a) and the models that currently dominate the literature are individual-based theories, such as Endsley's *information processing-based three-level model* (Endsley, 1995a), Smith and Hancock's *perceptual cycle model* (Smith and Hancock, 1995) and Bedny and Meister's (1999) *theory of activity* model. These three models are presented in Figure 1 and compared and contrasted in Table 1.

Endsley's model is generic and presents an intuitive description of SA, but its greatest utility lies in its simplicity and also the division of SA into three hierarchical levels, which allows the construct to be measured easily and effectively, and also supports the abstraction of SA requirements and the development of specific training strategies and design guidelines to support the acquisition of SA. Indeed none of the other models support such endeavors so effectively. The three-level model, however, is not without its flaws. For example, Endsley distinguishes between the *product* of SA and the *processes* that are used to achieve it, suggesting that the two are separate. Her definition is therefore contradictory, since it refers to the '*perception of the elements*', the '*understanding of their meaning*', and the '*projection of future states*', all of which could be taken to be *processes* involved in the development of SA. Also,

Theory	Domain Applications	Theoretical Underpinning	Process	Composition	Maturity	Measure	Process or Product?	Main Strengths	Main Weaknesses
Three Level Model (Endsley, 1995a)	Military (Aviation, Infantry Ops), Air Traffic Control, Aviation (Flight & Maintenance), Driving, Nuclear Power,	Information Processing Theory Recognition Primed Decision Making Model (Klein, 1990)	Perception of elements Comprehension of meaning Projection of future states	Perception, comprehension and projection of SA elements	18 years	SAGAT (Endsley, 1995b) SA Requirements Analysis (Endsley, 1993)	Product	1. Simple intuitive description of SA 2. Division of SA into levels is neat and permits simplistic measurement using SAGAT 3. Holistic approach that considers factors such as system & interface design, workload and training	1. Fails to cater for the dynamic nature of SA 2. SA process oriented definition is contradictory to the description of SA as a 'product' comprising three levels 3. Based on ill defined and poorly understood psychological models (e.g. information processing, mental models)
Perceptual Cycle Model (Smith & Hancock, 1995)	Air Traffic Control	Perceptual Cycle Model (Niesser, 1976)	Schema driven exploration & modification	Externally directed consciousness	11 years	Task Performance Risk Space (Smith & Hancock, 1995)	Process & Product	1. Dynamic description of SA acquisition, maintenance and update of schema 2. Sound theoretical underpinning 3. Completeness of model is attractive i.e. it describes both the process of acquiring SA and the product of SA	1. Does not translate easily to SA description and measurement 2. Limited applications 3. The actual correlation between SA and performance is complex and not yet fully understood
Theory of Activity Model (Bedny & Meister, 1999)	None	Theory of Activity Model (Bedny & Meister, 1999)	Orientational Stage Executive Stage Evaluative Stage	Incoming Information Goals Conceptual Model of Situation Past Experience Environmental Features Motivation towards task goals Subjectively relevant Task Conditions	7 years	N/A	Process & Product	1. Model offers a more complete, dynamic description of SA than the three-level model 2. Clear description of each functional blocks role in SA acquisition and maintenance is useful 3. Sound theoretical underpinning	1. Very limited application and the model lacks supporting empirical evidence 2. No measurement approach suggested 3. Individual approach that does not attempt to describe team SA

Theory	Domain Applications	Theoretical Underpinning	Process	Composition	Maturity	Measure	Process or Product?	Main Strengths	Main Weaknesses
Team SA Model (Endsley & Robertson, 2000)	Military, Aviation Maintenance	Three Level SA Model (Endsley, 1995a)	Perception of elements Comprehension of meaning Projection of future states Sharing of mental models	Individual SA Shared SA (Overlapping SA Requirements)	5 years	SAGAT (Endsley, 1995b) SA Requirements Analysis (Endsley, 1993)	Product	1. Extension of the popular and widely applied three-level model - sound theoretical underpinning and lots of supporting literature 2. Widely applied in a variety of domains 3. Comes with prescribed SA measurement approach (SAGAT)	1. More of a simplistic extension of the individual three level model than a team model in its own right 2. Measurement is complex and impractical for real-world distributed tasks
Team SA Model (Salas et al, 1995)	None	Teamwork Theory	Perception of elements Comprehension of meaning Projection of future states Team Processes	Individual SA Team Processes Information Seeking Information Processing Information Sharing	11 years	Individual SA Team Processes Compatibility of mental models TARGETS (Fowlkes et al, 1992)	Process & Product	1. Provides an insight into the team processes linked to team SA 2. Based on a review of teamwork literature 3. Relates model to team training and speculates on what to measure and how to measure it during team SA assessments	1. Measurement approach is more suited to assessing team behaviour and performance than SA and team SA measurement applications are scarce 2. The model is based on a review of the team literature rather than naturalistic or empirical study 3. Focussed more on team processes than on team SA
Team SA Model (Wellens, 1993)	Military	N/A	Individual SA Sharing of SA Team Processes	Overlapping SA Common Perspective	13 years	CITIES (Wellens, 1993) Post Task Questionnaire Task Performance	Process & Product	1. CITIES experimental paradigm developed specifically for assessing team SA	1. SA assessments restricted to CITIES VR environment
Distributed Cognition Approach (Artman & Garbis, 1998)	Teleoperations	Distributed Cognition Theory (Hutchins, 1995)	Shared & Distributed Models	Partly Shared and Partly Distributed Model of Situation	7 years	Observation/Field Study	Process & Product	1. Systems level description that permits both individual, collaborative and systemic SA assessments 2. Sound theoretical underpinning	1. Limited applications 2. No prescribed measurement approach 3. Does not describe individual SA processes
Mutual Awareness Team SA Model (Shu & Furuta, 2005)	Process Control	Three Level SA Model (Endsley, 1995a) Shared Co-operative Activity Theory (Bratman, 1992)	Individual SA Mutual Awareness	Individual SA Awareness of distributed context Awareness of team members and how they interact with context	1 year	TSA Simulation	Process & Product	1. Model attempts to describe the content of team SA and the behaviours involved in its development 2. Sound theoretical underpinning 3. Builds on existing SA theory and uses additional shared co-operative activity theory to present arguement	1. Complex description of team SA 2. Measurement approach is limited to authors domain 3. Limited application or validation
Distributed Situation Awareness Model (Stanton et al, 2006)	Military, Maritime, Energy Distribution, Aviation, Air Traffic Control, Emergency Services, Driving	Distributed Cognition Theory (Hutchins, 1995) Distributed SA Theory (Artman & Garbis, 1998)	Individual SA Sharing of Knowledge Elements Team Processes	System Level Emergent Property Activated Knowledge Shared Knowledge	1 year	Propositional Networks (Stanton, Salmon, Walker, Baber & Jenkins, 2005)	Process & Product	1. Systems level description that permits both individual, collaborative and systemic SA assessments 2. Sound theoretical underpinning 3. Has been applied in a variety of collaborative domains	1. DSA description and measurement is subjective and often occurs post-task 2. Propositional Networks methodology lacks validation 3. Does not describe individual SA processes

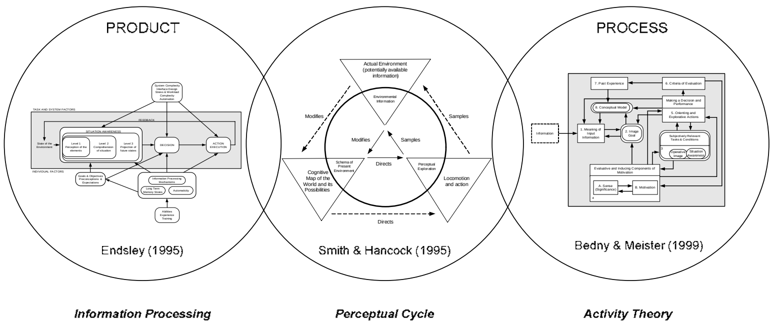


Fig. 1. Situation Awareness Models

Endsley has been criticized for using poorly understood approaches to underpin the model. Smith and Hancock (1995), for example, suggest that its reference to mental models, which themselves are ill-defined, is problematic. Similarly, Uhlarik & Comerford (2002) criticise Endsley's theory for its use of an information-processing model that contains ill-defined psychological constructs that are subject to great debate themselves. The model has also been criticised for its inability to cope with the dynamic nature of SA (Stanton, Chambers & Piggott, 2001; Uhlarik & Comerford, 2002).

The perceptual cycle model (Smith and Hancock, 1995) has sound underpinning theory (Neisser, 1976) and is complete in that it refers to both the *process* (continuous sampling of the environment) and the *product* (continually updated schema) of SA. Unlike the three-level model, Smith and Hancock's model therefore grasps the dynamic nature of the construct. On the downside, the perceptual cycle model has not received anywhere near the attention that Endsley's model has, and its description of SA makes measuring the construct complex. Bedny & Meister's description of the way in which SA dynamically modifies interaction with the world and then interaction with the world dynamically modifies SA is logical, and goes beyond the static perspective taken by Endsley's model. Despite this, there is a distinct lack of empirical evidence available to support their model, and consequently it has received less attention than Endsley's model.

In summary, each theory has its useful components. As Stanton et al (2001) point out, 'there appears to be an element of truth in all of them'. In terms of theoretical utility, Smith and Hancock's model is perhaps the most useful, since it caters for the dynamic aspects of SA (Stanton et al, 2001), whereas the theory of activity model is useful since it details an individuals internal activities involved when developing SA. Endsley's three level model, on the other hand, offers a very neat and intuitive description of SA which allows researchers to measure the construct simplistically and also to abstract SA requirements at each level. As a consequence of this, the three level model has been the most useful for researchers wishing to describe operator SA in complex systems and also for measuring the construct. Further, Endsley's work has proved to be the most useful for informing system design and evaluation (e.g. Endsley, Bolte & Jones, 2003). One consistent theme that links these theories is that they describe SA acquisition and maintenance from the point of view of an individual operator and so fail to cater for team SA or SA during collaborative activity.

4 Situation Awareness in Collaborative Systems

Salas, Prince, Baker and Shrestha (1995) point out that there is a lot more to team SA than merely combining individual team member SA. On a simple level we can say that team SA is multidimensional, comprising *individual* team member SA, *shared SA* between team members, and also the *combined SA* of the whole team. Add to this the complexity of the systems in which team SA is critical (e.g. military command and control systems, emergency service domains etc), the various team processes involved in the development and maintenance of team SA (e.g. communication, co-ordination etc) and the various factors affecting team SA levels, and the complexity of the construct quickly becomes apparent. Based on a review of the literature Salas et al (1995) proposed a framework of team SA, suggesting that it comprises two critical, but poorly understood, processes: *individual SA* and *team* processes. They subsequently defined team SA as “the shared understanding of a situation among team members at one point in time” (Salas et al, 1995). Other researchers have suggested that team SA refers to the level of overlap in SA between team members. Endsley (1989), for example, defines team SA as, “the degree to which every team member possesses the situation awareness required for his or her responsibilities” (Endsley, 1989) and shared SA as “the degree to which team members have the same SA on shared SA requirements” (Endsley & Jones, 2001). Endsley (1989) suggests that, during team activities, SA can overlap between team members, in that individuals need to perceive, comprehend and project SA elements that are specifically related to their specific role in the team, but also elements that are required by themselves and by members of the team. Successful team performance therefore requires that individual team members have good SA on their specific elements and also the same SA for those elements that are shared (Endsley & Robertson, 2000).

In conclusion to a review of team SA in aircraft maintenance teams, Endsley and Robertson (2000) suggested that good team SA is dependent on team members understanding the meaning of the information that is passed between one another. According to Endsley and Robertson (2000) this means that teams need to share pertinent data and the higher levels of SA, such as the significance of SA elements to the team’s goals and also projected states. Endsley and Robertson (2000) go on to suggest that the primary factors linked to team performance are shared goals, the interdependence of team member actions, and the division of labour between team members. This means that some SA requirements are independent but also that team members possess shared goals and perform interdependent activities, which means that they also possess shared SA requirements (Endsley and Robertson (2000). Wellens (1993) suggests that the key to team SA lies in the arrangement of teams so that sufficient overlap between team member SA occurs to support co-ordination, but also so that sufficient separation between members allows individual SA acquisition. Wellens (1993) defines group or team SA as “the sharing of a common perspective between two or more individuals regarding current environmental events, their meaning and projected future”.

In terms of how team SA is acquired and maintained, Bolstad & Endsley (2000) suggested that the development of shared SA involves the following four factors: *shared SA requirements* (e.g. the degree to which team members understand which

information is needed by other team members), *shared SA devices* (e.g. communications, shared displays and the shared environment), *shared SA mechanisms* (e.g. shared mental models), and *shared SA processes* (effective team processes for sharing relevant information). One of the key concepts thought to be critical to team SA is the notion of shared mental models. Fiore, Cuevas & Salas (2003) suggest that a shared mental model is, "the activation in working memory of team and task-related knowledge while engaged in team interaction". It is thought that shared mental models facilitate communications between team members (Perla et al, 2000) and can allow team members to forecast the behaviour of other team members (Salas, Stout and Cannon Bowers, 1994). Endsley (1995a) even argues that team SA is more reliant on mental models than it is on verbal communication.

The role of team behaviours, such as team communication, co-ordination, adaptability, and co-operation and team attitudes, such as team trust, collective efficacy and orientation (Fiore et al, 2003) is often neglected when describing team SA. It seems logical to assume that an increased level of teamwork will lead to enhanced levels of team SA, however the specific relationship between team SA and team behaviours and attitudes remains largely unexplained. Most researchers have focused on communication as the key element in the acquisition of team SA. Nofi (2000) for example, singles out *communication* is the most critical element in the creation of team SA, whilst Entin and Entin (2000) suggest that communication is a prerequisite for high levels of team SA. Salas et al (1995) also highlighted the importance of communication in team SA acquisition.

The team SA theories identified in the literature are compared and contrasted in Table 2. The review leads us to conclude that there is a lack of a unified, universally accepted model of team SA. The approaches presented focus on either a shared awareness of the situation, sometimes called a 'common understanding' or on the overlap between team member SA requirements. It seems that there is a lack of a model that fully describes the processes involved, the content of team SA and also the factors impacting team SA. Based on a synthesis of the literature, it can be concluded that team SA comprises a team's compatible awareness of the situation. Team members must possess SA-related to their individual roles and goals within the team (some of which may be common or 'shared' with other team members), whilst also holding SA-related to other team members, including an awareness of other team members activities, roles, and responsibilities, and also to the team overall, including goals and performance. SA-related data and knowledge is distributed around the team through team processes such as communication, co-ordination and collaboration, and serves to inform and modify team member SA, which is also informed and modified by the overall teams SA. Thus a tripartite composition of team SA is apparent: *individual team member SA* (some of which may be common or 'shared' with other team members), *SA of other team members*, and *SA of the overall team*. This simplistic view of team SA is presented in Figure 2.

The team or shared SA theory perspective is limited for a number of reasons (some of which are not alluded to here due to space limitations). Take, for example, military Networked Enabled Capability (NEC) scenarios. Such tasks involve numerous agents and artifacts working both collaboratively and in isolation from one another whilst being dispersed geographically, often over great distances. Viewing and assessing team SA in such environments is acutely complex. The dispersed, real world nature of

tasks in these environments inhibits (or at least makes impractical) the use of objective probe techniques such as the Situation Awareness Global Assessment Technique (SAGAT; Endsley, 1995b). Also, the simplistic description of individual SA and shared SA can be viewed differently. For example, it may be that team members do not need to have SA of the overall team's situation, and that they only require SA for their specific task and role within the team, SA that is different but compatible with their counterparts. Therefore to suggest that all team members have their own SA and also shared SA with team members and of the overall team could be an oversimplification. Any cog in a machine does not need to 'know about' all of the other cogs, but it does need to be able to connect with adjacent cogs – thus we are suggesting that 'compatibility' is the key to team SA, rather than 'sharedness'. Any sharing of goals, intent, and understanding arises out of the need of the individual agents to perform their tasks and not for its own sake. We feel that the ideas of 'sharing' have mutated into a vague belief that sharing ensures a cohesive team, but we argue that 'compatibility' leads to cohesiveness.

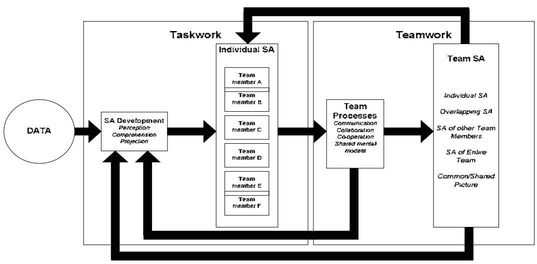


Fig. 2. Team SA

5 Distributed Situation Awareness

A more recent theme to emerge is the concept of distributed or systemic SA. Based upon distributed cognition theory (Hutchins, 1995), DSA approaches view SA as an emergent property (i.e. relationship between system elements) of collaborative systems. DSA is achieved through co-ordination between system units (Artman & Garbis, 1998) and is viewed systemically rather than as an individual endeavor or shared understanding. Whilst recognising that individuals possess their own SA for a particular situation and that they share this understanding of the situation (Artman & Garbis, 1998), DSA approaches assume that collaborative systems posses cognitive properties (such as SA) that are not part of individual cognition. No single member has the overall SA, rather it is distributed around the system. Artman and Garbis (1998) define team SA as,

> "The active construction of a model of a situation partly shared and partly distributed between two or more agents, from which one can anticipate important future states in the near future" (Artman & Garbis, 1998)

Stanton, Stewart, Harris, Houghton, Baber, McMaster, Salmon, Hoyle, Walker, Young, Linsell, Dymott & Green (2006) recently proposed a theory of DSA for complex systems. Stanton et al suggest that SA-related knowledge is distributed across the agents and artefacts comprising the system and that these knowledge 'themes' or 'topics', labelled *knowledge elements,* represent what agents 'need to know' in order to achieve success during task performance. In this case, the term knowledge represents the relationship between concepts (Shadbolt & Burton, 1995) and refers to task-level knowledge, which relates to the goals and sub-goals of the task being performed. The ownership, usage and sharing of knowledge is dynamic and is dependent upon the task and its associated goals. Agents therefore have different SA for the same situation, but their SA can be overlapping, compatible and complementary, and deficiencies in one agents SA can be compensated by another agent. Stanton et al define DSA as 'activated knowledge for a specific task, at a specific time within a system'.

Stanton et al (2006) propose then, that a situation requires the use of appropriate knowledge (held by individuals, captured by devices etc.) that relates to the state of the environment and those changes as the situation develops. The 'ownership' of this knowledge is initially at the system, rather than individual level. This notion could be further extend to include 'meta-SA', where its knowledge of other agents' knowledge is contained in the system, such that each agent could potentially know where to go when they need to find something out. It is important to note that DSA approaches do not contend that individual SA perspectives are redundant; rather they provide an alternative, but complementary approach to viewing and describing SA in collaborative systems. For example, in extending the DSA to Endsley's three level model approach, we assume that within collaborative systems, some individuals are engaged in perception tasks, some are engaged in comprehension and in the projection tasks and others are engaged in response execution tasks.

6 Discussion

In this article we have attempted to present a synthesis of the literature on SA and team SA. The review indicated that the prominent SA models focus on SA acquisition and maintenance largely from the point of view of individual operators in complex systems. As a result of this, team SA as a theoretical construct remains a challenge to the human factors community. Team SA, at a very simplistic level, comprises individual team member SA, some of which may be common or shared, SA of other team members, and the holistic SA of the team as a whole. The development and maintenance of team SA is impacted by a myriad of factors, both at to the individual level (e.g. experience, stress, workload etc), team level (e.g. team behaviours such as communication and collaboration) and the systems level (e.g. system design, procedures, technology used etc). This multidimensional composition of team SA ensures that describing and assessing SA during real world collaborative tasks is complex. Despite the successful extension of Endsley's three level model to the team environment (Endsley and Robertson, 2000, Endsley and Jones, 2001), one still feels that, theoretically at least, team SA requires further investigation. Based on the review, it is concluded that recently emerged DSA approaches (e.g. Stanton et al,

2006) are suited to describing and assessing SA in real world collaborative environments, or in environments where SA elements are not well defined a priori. Viewing SA in this way is fruitful for a number of reasons, including that it permits a systemic description of the knowledge comprising SA (which can be extrapolated to an individual SA level) and it allows judgements to be made on potential barriers to SA acquisition and maintenance. Further, considering SA in this way ensures that team SA is viewed in its entirety, rather than as its component parts (i.e. individual team member SA). In collaborative systems, tasks are rarely performed entirely independently of others, especially in complex situations and when critical decision-making is required (Artman & Garbis, 1998) – these activities tend to require coordinated activity between several individuals (Cannon-Bowers and Salas, 1990; cited in Salas, Prince, Baker and Shrestha, 1995). It is important therefore that team SA assessments consider this co-ordination.

References

Artman, H., Garbis, C.: Situation awareness as distributed cognition. In: Proceedings of ECCE' 98, Limerick (1998)

Bedny, G., Meister, D.: Theory of activity and situation awareness. International Journal of Cognitive Ergonomics 3(1), 63–72 (1999)

Bolstad, C.A., Endsley, M.R.: The effect of task load and shared displays on team situation awareness. In: Proceedings of the 14th Triennial Congress of the International Ergonomics Association and the 44th Annual Meeting of the Human Factors and Ergonomics Society (2000)

Endsley, M.R.: Final report: Situation awareness in an advanced strategic mission (NOR DOC 89-32). Northrop Corporation, Hawthorne, CA (1989)

Endsley, M.R.: Towards a theory of Situation Awareness in Dynamic Systems. Human Factors 37, 32–64 (1995)

Endsley, M.R.: Measurement of Situation Awareness in Dynamic Systems. Human Factors 37, 65–84 (1995)

Endsley, M.R., Jones, W.M.: A model of inter- and intrateam situation awareness: Implications for design, training and measurement. In: McNeese, M., Salas, E., Endsley, M. (eds.) New trends in cooperative activities: Understanding system dynamics in complex environments, Human Factors and Ergonomics Society, Santa Monica, CA (2001)

Endsley, M.R., Robertson, M.M.: Situation awareness in aircraft maintenance teams. International Journal of Industrial Ergonomics 26, 301–325 (2000)

Endsley, M.R., Bolte, B., Jones, D.G.: Designing for situation awareness: An approach to human-centred design. Taylor and Francis, London (2003)

Entin, E., Entin, E.: Assessing Team Situation Awareness in Simulated Military Missions. Ergonomics for the New Millennium. In: Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society, San Diego, California. vol. 1, pp. 73–76 (2000)

Fiore, S.M., Salas, E., Cuevas, H.M., Bowers, C.A.: Distributed coordination space: toward a theory of distributed team process and performance. Theoretical Issues in Ergonomics Science 4, 340–364 (2003)

Fracker, M.: Measures of Situation Awareness: Review and Future Directions (Rep. No.AL-TR-1991-0128). Wright Patterson Air Force Base, Ohio: Armstrong Laboratories, Crew Systems Directorate. (1991)

Hutchins, E.: Cognition in the wild. MIT Press, Cambridge Massachusetts (1995)
Niesser, U.: Cognition and reality: Principles and implications of cognitive psychology. Freeman, San Francisco (1976)
Nofi, A.: Defining and Measuring Shared Situational Awareness, DARPA (2000), http://www.thoughtlink.com/publications/DefiningSSA00Abstract.htm.
Perla, P., Markowitz, M., Nofi, A., Weuve, C., Loughran, J., Stahl, M.: Gaming and Shared Situation Awareness. DARPA GamingSSA00Abstract.htm (2000), http://www.thoughtlink.com/publications/
Salas, E., Stout, R.J., Cannon-Bowers, J.A.: The role of shared mental models in developing shared situational awareness. In: Gilson, D., Garland, D.J., Koonce, J.M. (eds.) Situational Awareness in Complex Systems: proceedings of a CAHFA conference, Embry-Riddle Aeronautical University Press, Daytona Beach, FL (1994)
Salas, E., Prince, C., Baker, P.D., Shrestha, L.: Situation awareness in team performance. Human Factors 37(1), 123–126 (1995)
Shu, Y., Furuta, K.: An inference method of team situation awareness based on mutual awareness. Cognition Technology & Work. 7, 272–287 (2005)
Shadbolt, N.R., Burton, M.: Knowledge elicitation: A systemic approach. In: Wilson, J.R., Corlett, E.N. (eds.) Evaluation of Human Work: A Practical Ergonomics Methodology, pp. 406–440 (1995)
Smith, K., Hancock, P.A.: Situation awareness is adaptive, externally directed consciousness. Human Factors 37(1), 137–148 (1995)
Stanton, N.A., Chambers, P.R.G., Piggott, J.: Situational Awareness and safety. Safety Science 39, 189–204 (2001)
Stanton, N.A., Stewart, R., Harris, D., Houghton, R.J., Baber, C., McMaster, R., Salmon, P., Hoyle, G., Walker, G.H., Young, M.S., Linsell, M., Dymott, R. Green, D.A.: Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. Ergonomics.49, 12-13, 1288–1311 (2006)
Uhlarik, J., Comerford, D.A.: A review of situation awareness literature relevant to pilot surveillance functions (DOT/FAA/AM-02/3). Department of Transportation. Federal Aviation Administration, Washington, DC, U.S (2002)
Wellens, A.R.: Group situation awareness and distributed decision making: From military to civilian applications. In: Castellan, N.J. (ed.) Individual and Group Decision Making: Current Issues, pp. 267–287. Erlbaum Associates, Mahwah (1993)

Stress and Managers Performance: Age-Related Changes in Psychophysiological Reactions to Cognitive Load

Sergei A. Schapkin[1], Gabriele Freude[1], Udo Erdmann[1], and Heinz Ruediger[2]

[1] Federal Institute for Occupational Safety and Health, Berlin, Germany
[2] Technical University of Dresden, Dresden, Germany

Abstract. Work ability of elderly managers may decrease probably due to decreased cognitive flexibility. Moreover, cardiovascular reactivity to changing task demands might be less efficient in elderly. Younger (36-45) and elderly (49-60) German managers had to perform a switching task when they had to switch continuously between different rules of information processing or to use the same rule. Although the performance did not differ between groups, the SBP was higher and the HRV was lower in elderly. In addition, elderly showed an increased rigidity of cardiovascular functioning against changing task demands.

1 Introduction

Managers have been confronting with complex and continuously changing tasks requiring cognitive flexibility and effort to maintain a constant level of job performance during a long time period. However, these professionally critical aspects of manager's job were not sufficiently investigated up to now. Performance of a cognitive task requires effort which can be either computational or compensatory [1]. The computational effort is directly related to task complexity in terms of resources allocating to a task under normal conditions, when there is no need to set new goals and adopt new working strategies. On the contrary, the compensatory effort becomes necessarily when an individual is acting under suboptimal conditions (i.e. stress, fatigue etc.) and trying to keep a constant performance level. Both effort types may be distinguished by cardiovascular changes that have been considered as reliable indicators of mental workload and effort [2]. The performance of a cognitive task as compared to rest is usually accompanied by an increase of heart rate (HR), blood pressure (BP) and a decrease of heart rate variability (HRV). The more complex the task (e.g. the more mental operations it requires) the greater the changes of HR, BP and HRV. Decrease of task complexity lead to opposite changes of these parameters. Such flexible adaptation of the cardiovascular system to task demands may be seen as physiological index of computational effort that typically accompanied by high and stable performance level. In case the cardiovascular parameters and performance dissociate a suboptimal regulation may be assumed when a compensatory effort comes into play to prevent performance decrement. Another index of this effort type would be a rigidity of cardiovascular functioning resulting in reduced variation of cardiovascular parameters with task complexity; e.g. when HRV remains decreased in an easy task or when a slow post-task recovery of HRV and HR is observed.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 417–425, 2007.

Incremental ageing of work population in Western Europe during last two decades has been resulted in a prolonged employment among managers as well. However, systematic studies of cardiovascular reactions to mental workload in elderly workers are still scarce and, to our knowledge, absent in managers. Executive control processes underlying cognitive flexibility have been found to be quite susceptible to age [3]. Additionally, the age-related dysregulation of cardiovascular functions has been usually observed [4]. Therefore, resources to perform an “executive” task are assumed to be shortened in elderly who should exert a compensatory effort to maintain performance. Keeping this in mind, one might expect the performance decrement as well as inefficient cardiovascular adaptation to changing task demands in elderly manager.

The task switching paradigm [5] belongs to a family of “executive” tasks and can be used to investigate cognitive flexibility that is critical for manager's job. During the task participants have to change the processing rules when the information context is changed. The change results usually in a lengthening of reaction time and worsening of accuracy usually defined as switching costs (SK). The global SK are computed as difference between performance in a task containing an unique processing rule (single task) and a task when the switching of rules is required (switch task). The local SK are computed within the switch task as the difference between trials when the processing rule remains the same (repeat trials) and trials when the rule has to be changed (switch trials). The task switching involves control mechanisms providing the selection of a relevant task set and the inhibition of an irrelevant ones as well as continuous updating of current task requirements in working memory. These processes are thought to be a part of executive functioning that is sub-served by the medial and dorsolateral prefrontal cortex of the brain [6]. Based on the previous data that demonstrated larger SK in elderly than younger adults [7], [8] we expected to replicate the effect in the present study. Alternatively, elderly managers might be more experienced in handling with alternating tasks and thereby may outperform their younger colleagues.

In sum, in the present study switch tasks differed in complexity were imposed and cardiovascular parameters were registered to examine the psychological and physiological adaptation of younger and elderly managers to changing task demands.

2 Method

2.1 Participants

Participants in the present study were seventeen German managers of a steel factory. They were divided by median-split (47 years) into younger (8 men, 36-45 years) and elderly (8 men, 49-60 years) groups while the participant with the median age was excluded from the analysis. Participants were right-handed, had normal hearing, normal or corrected vision and no heart diseases. The participation in the study was voluntary, as the study was presented as a part of a routine clinical examination. All participants gave their informed consent for participation.

2.2 Task

The switch task employed was modified from the alternating-run paradigm [5] participants had to alternate regularly every two trials between task A and B (i.e. AABBAABB etc.). The participants had to compare digits 2, 3, 4, 6, 7, 8 with the cue (digit “5”) either numerically (task A) or by font size (task B). If the cue appeared in white, the participants had to decide whether the following digit was numerically less or more than 5 (left response key – less than 5, right response key - more than 5). If the cue appeared in black, the comparison by font size was required (left response key - smaller than 5, right response key – larger than 5). Each trial started with a white fixation cross (15 x 15 mm) remained in the centre of the screen for 1000 ms followed by a digit “5” (8mm width x 15mm high) for 300 ms depicted in white or black and framed in green circle (diam. 25 mm). After cue off-set the cue-stimulus interval (CSI) of 500 ms was presented till an imperative stimulus appeared for 200 ms. The imperative stimuli were digits 2, 3, 4, 6, 7, 8, presented in small (7 x 10 mm) or large (12 x 18 mm) font size and equally often in each task. The distance to the screen was 70 cm. One half of the imperative stimuli were compatible and another half was incompatible with the response. On compatible trials attributes of both tasks were mapped onto the same response key, while on incompatible trials onto the different response keys. The reaction time was measured within 4000 ms from the onset of an imperative stimulus. If the correct response was made within the response window, next trial was started. The false response was followed by a delay of 200 ms and a 500 Hz tone of 100 ms duration. Omissions were indicated by a 3000 Hz tone of 100 ms duration presented immediately after the end of the response window. In addition to the switch task, the task A (numeric comparison) and task B (size comparison) were presented separately while the same timing features were used in both single tasks. The single tasks consisted of 48 trials and the switch task consisted of 192 trials (50% switch and 50% repeat trials).

2.3 Procedure

Participants were tested individually in a clinical examination room of the steel factory. The experiment began with the measure of the BD and HR via upper arm cuff followed by 3 min beat-to-beat BD registration in rest condition. Thereafter participants were presented three training tasks: task A (12 trials), task B (12 trials) and, finally, switch task (32 trials). The sequence of training tasks was equal for all participants. In the main session the task sequence was balanced across participants. The beat-to-beat BD was registered throughout experiment. Finally, the post-task beat-to-beat BD during 3 min was registered followed by rest measure of the BD and HR via upper arm cuff.

2.4 Cardiovascular Measures

At the start and the end of the experiment the BD and HR were registered via cuff from left upper arm by an electronic device Boso Medicus (Bosch+Sohn LTD, Germany). During the experiment beat-to-beat BP was registered continuously from left middle finger using a Finapres device (Ohmeda, USA). The RR interval was computed off-line as the interval between two successive BP maxima. As a HRV

measure the RMSSD index was computed. Data were processed by the trigonometric regressive spectral analysis (TRS, [9]) based on the Cosinor method and can be used to beat-to-beat recordings of cardio-vascular variables like RR interval, systolic blood pressure, stroke volume etc. The advantage of such analysis is that data can be assessed in non-equidistant steps (in contrast to Fourier transformation) and it also minimises practical problems occurring due to insufficient frequency resolution, aliasing and can assess data segments of varying length. TRS detects underlying rhythms using the following equation: $F = \sum(\text{variable}(t_i) - \text{Reg}(t_i))^2 \Rightarrow \text{Minimum}$. In this function, non-equidistant data (i.e. RR interval, systolic or diastolic blood pressure and others) is used and parameters of amplitude (a), phase shift (φ) and frequency (ω) are represented as a trigonometric function in which $\text{Reg}(ti) = a*\sin(\omega ti+\varphi i)$. Using regression analysis, amplitude, frequency and phase shift can now be estimated using partial differential quotients (i.e., $\delta F/\delta a$; $\delta F/\delta \omega$; and, $\delta F/\delta \varphi$).

2.5 Data Analysis

Each response was scored as correct if the appropriate key had been pressed in a time window lasting from 200 to 4000 ms after stimulus onset. The 200 ms limit was chosen to avoid counting any premature responses.

To analyse the global switch costs the error rate (% Err) and the reaction time (RT) for correct responses in switch task were compared with those in single tasks. They were subjected to an ANOVA with Task (task A, task B, switch) as a within-subject factor and Age (younger vs. elderly) as a between-subject factor. To analyse the local switch costs %Err and RT for correct responses for switch and repeat trials separately were subjected to an ANOVA with Trial (switch vs. repeat) as a within-subject factor and Age (younger vs. elderly) as a between-subject factor.

To examine how cardiovascular parameters varied with task demands SBP, DBP, HR, HRV, were subjected to ANOVAs with Condition (pre-test, task A, task B, switch task, post-test) as a within-subject factor and Age (younger vs. elderly) as a between-subject factor. Additional ANOVAs with the same within-subject factor was performed for younger and elderly group separately in order to test Condition effect within each group. The SBP and DBP measured via upper arm cuff were analyzed by ANOVA with Condition (pre-test, post-test) as a within-subject factor and Age (younger vs. elderly) as a between-subject factor.

Greenhouse-Geisser epsilon was applied to correct a possible lack of sphericity Simple contrasts were used to clarify effects across factor levels. ANOVAs were performed by SPSS for Windows 14.0.

3 Results

3.1 Behavioural Data

Enhanced global switch cost in terms of longer RT and higher %Err were obtained in switch as compared to single tasks (Task, RT: ($F(2,28) = 15.982$, $p = .001$; Task, %Err: ($F(2,28) = 33.88$, $p < .000$). Analysis of contrasts revealed the longer RT and higher %Err in the switch task than in task A while both performance measures were

the lowest in the task B. A marginally significant effect of Age (F(2,28) = 4.055, p = .064) was due to higher %Err in elderly than younger. Post-hoc tests revealed higher %Err in elderly than in younger in switch task and in task A (numeric comparison) only. No significant effects on local switch costs in terms of RT or %Err were found.

3.2 Cardiovascular Data

A significant main effect of Condition (F(4,56) = 8.887, p = .001) indicated an enhanced HR in switch task as compared to both single tasks (Fig. 1).

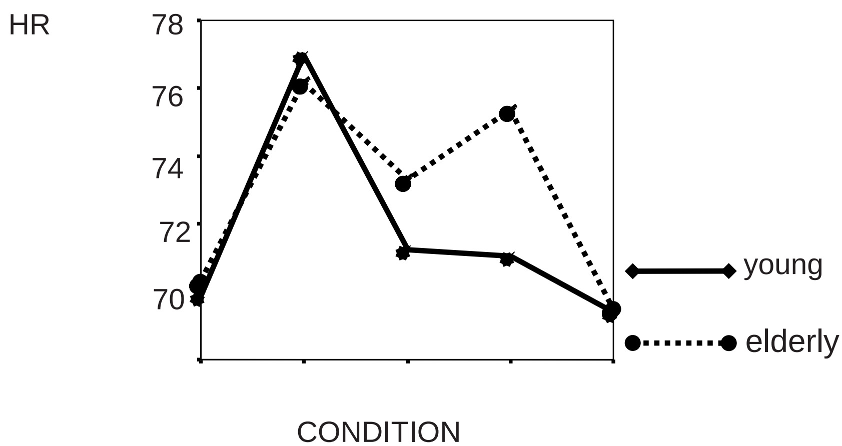


Fig. 1. Heart rate (HR, beat per min) as a function of experimental condition and age: 1-pre-task rest, 2- switch task, 3-task A (numeric), 4- task B (size), 5- post-task rest

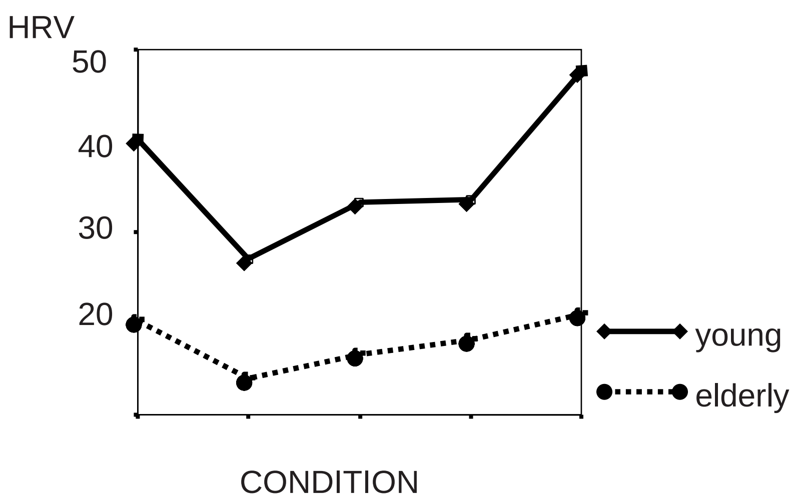


Fig. 2. Heart rate variability (HRV) as a function of experimental condition and age: 1-pre-task rest, 2- switch task, 3-task A (numeric), 4- task B (size), 5- post-task rest

In turn, the HR in single tasks was higher than in rest conditions and did not differ between single tasks. An Age by Condition interaction was non-significant. However, on the Fig.1 a tendency to increased HR in elderly in task B was clearly seen. The within-subject contrasts of switch task against task B revealed a significant Age by Condition interaction ($F(1,14) = 6.189$, $p = .014$) that was due to increase of HR in elderly in the task B only.

A significant main effect of Condition ($F(4,56) = 10.351$, $p<.000$) was attributed to a reduced HRV in the switch task as compared to both single tasks in which the HRV was lower as in rest conditions; the HRV did not differ between single tasks (Fig.2). An Age by Condition interaction ($F(1,14) = 2.25$, $p = .051$) was due to the fact that HRV varied with experimental conditions in younger group to a greater extent than in elderly one. Separate ANOVAS revealed a clear-cut Condition effect in younger (Task: $F(4,28) = 9.875$, $p = .004$) but not in elderly (Task: $F(4,28) = 1.81$, $p = .216$) group.

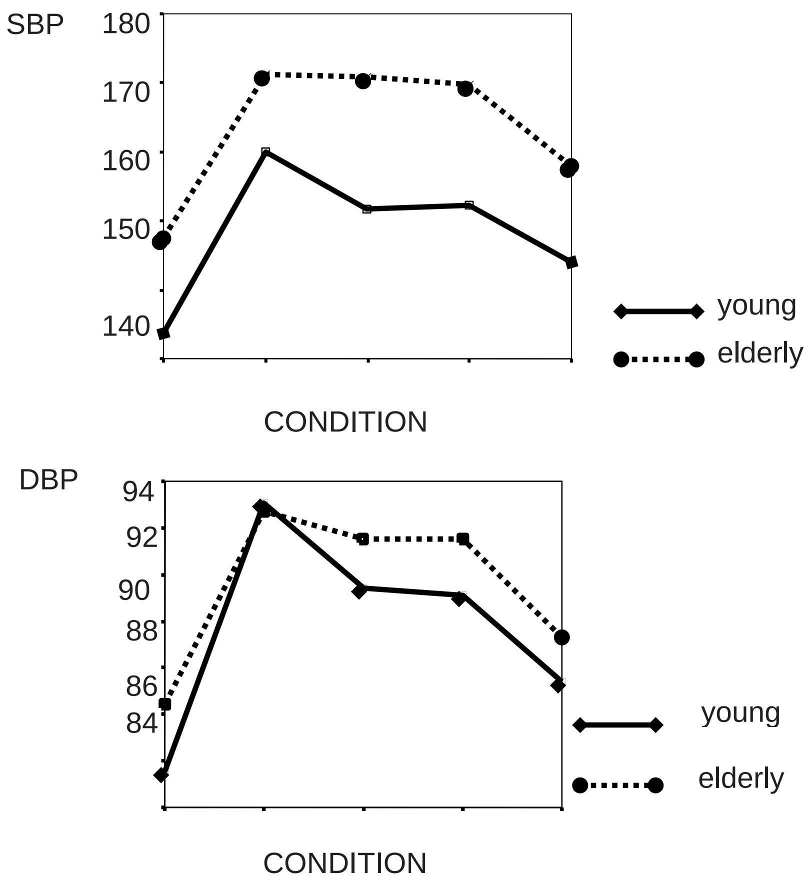


Fig. 3. Systolic (SBP, mmHg) and diastolic (DBP, mmHg) blood pressure and as a function of experimental condition and age: 1-pre-task rest, 2- switch task, 3-task A (numeric), 4- task B (size), 5- post-task rest

Both SPB and DBP were higher during three cognitive task than in both rest conditions while no difference among three tasks was obtained (SBP: $F(4,56) = 20.579$, $p<.000$; DBP: $F(4,56) = 20.579$, $p<.000$, Fig. 3). The SBP was generally higher in elderly (Age: $F(1,14) = 6.337$, $p = .025$).

The HR measured via upper arm cuff revealed a marginally significant Condition by Age interaction ($F(1,14)=4.047$, $p=.064$, Fig. 4). Younger had higher HR than elderly at the start, whereas the opposite effect at the end was seen (Fig.4).

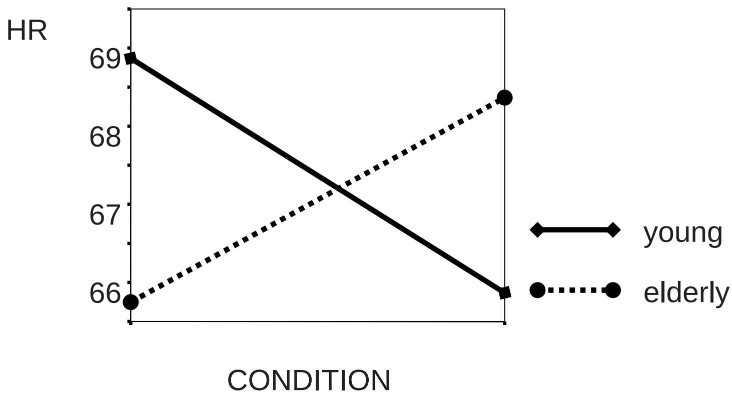


Fig. 4. Heart rate (beat per min) measured by common method as a function of experimental condition and age: 1-pre-task rest, 2-post-task rest

4 Discussion

In the present study age-related decrease of cognitive flexibility and physiological adaptation to changing task demands among managers were assumed. In the switch task we did not find age differences in reaction time, while only moderate accuracy decrement in elderly was seen. Two possible explanations may be taken into consideration. First, elderly managers are more experienced to handle with alternating tasks. If it would be true, they would show reduced physiological costs. In contrast, the physiological costs in elderly were higher than in younger suggesting that former prevented performance decrement by compensatory effort [1]. As we assumed, this effort type is related to a reduced responsivity of cardiovascular system to alternating tasks demands. In fact, elderly managers demonstrated a higher rigidity of cardiac responses to varied task complexity that was clearly seen on HRV. The HRV represents a mechanism providing allocation of physiological resources to a task in terms of changing blood supply of organs involved in task performance. Thus, the HRV should be the highest in rest condition, a somewhat reduced in an easy task, and be the lowest in a difficult task. This kind of profile was obtained in younger group indicating an efficient adaptation of their cardiovascular function to experimental conditions (Fig 2). In addition, a tendency to higher post-test than pre-test HRV in younger was seen suggesting a quick cardiac recovery in this group. On the contrary, the HRV in elderly group was insusceptible to alternating conditions. The effect is

generally in line with data has been demonstrated a reduced heart rate response to cognitive workload at older ages [4], [10]. However, some authors did not find age differences in HRV response to cognitive load [11]. As the authors used a simple reaction time task, the result is undoubtedly due to task simplicity, whereas a complex cognitive task in our study was imposed by which a clear-cut age-related HRV differences were provoked.

Several studies have shown that systolic blood pressure (SBP) increases with age [12]. We also obtained a higher SBP in elderly than in younger group (Fig. 3) that may be due to a reduced vagal control of the BP with advancing age [13] and may probably lead to hypertension in elderly. The DBP did not differ between groups suggesting that mechanisms regulating DBP is less vulnerable to age. Higher HR in the pre- than in post-test conditions observed in elderly may be interpreted as increased reaction to experimental situation (Fig. 4). In younger group the pre-test increase of HR may be attributed to initial resource mobilization, whereas the post-test decrease of HR to quick recovery of cardiac function when the experiment was finished. The decreased responsivity of cardiovascular system to task demands as well as its delayed post-test recovery in elderly may increase risks for heart diseases [14]. In turn, cardiac dysfunction (e.g. hypertension) may lead to cognitive impairments, especially in “executive” tasks [15], [16]. It has been supposed that the damage of vessels surrounding white matter of the brain may contribute to deficits in executive functions in hypertensives [17].

Finally, some limitations of the study should be mentioned. Age difference between “elderly” and “younger” participants was very small as compared with common aging studies usually investigating the contrast age groups. It may have resulted in the attenuating of age-related effects on performance and cardiac responses. Further investigations using more contrast groups are required to pinpoint aging effects and to assess possible health risks related to a decreased responsivity and delayed recovery of cardiovascular system in elderly managers.

References

1. Hockey, G.R.J.: Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. Biol. Psychol 45, 73–93 (1997)
2. Mulder, L.: Measurement and Analysis methods of heart rate and respiration in applied environment. Biol. Psychol 43, 205–236 (1992)
3. West, R.L.: An application of prefrontal cortex theory to cognitive aging. Psychological Bul. 120, 272–292 (1996)
4. Steptoe, A., Fieldman, G., Evans, O., Perry, L.: Cardiovascular risk and responsivity to mental stress: the influence of age, gender and risk factors. J. Card. Risk 3, 83–93 (1996)
5. Rogers, R.D., Monsell, S.: Costs of a predictable switch between simple cognitive tasks. J.Exp. Psychol.: Gen. 124, 207–231 (1995)
6. DiGirolamo, G.J., Kramer, A.F., Barad, V., Cepeda, N., Weissman, D.H., Wszalek, T.M., Cohen, N.J., Banich, M., Webb, A., Beloposky, A.: General and task-specific frontal lobe recruitment in older adults during executive processes: A fMRI investigation of task switching. Neurorep 12, 2065–2072 (2001)

7. Kramer, A.F, Hahn, S., Gopher, D.: Task coordination and aging: Explorations of executive control processes in the task switching paradigm. Acta. Psychol 101, 339–378 (1999)
8. Hillman, C.H., Kramer, A.F., Belopolsky, A.V., Smith, D.P.: A cross-sectional examination of age and physical activity on performance and event-related brain potentials in a task switching paradigm. Int. J. Psychophysiol 59, 30–39 (2006)
9. Ruediger, H., Klinghammer, L., Scheuch, K.: The trigonometric regressive spectral analysis – a method for mapping of beat/to/beat recorded cardiovascular parameters on to frequency domain compared with Fourier transformation. Com. Meth. Prog. Biomed. 58, 1–15 (1999)
10. Garwood, M., Engel, B.T., Capriotti, R.: Autonomic nervous system function and aging: response specificity. Psychophysiol 19, 378–385 (1982)
11. Wood, R., Maraj, B., Lee, C.M., Reyes, R.: Short-term heart rate variability during a cognitive challenge in young and older adults. Age. and Ageing 31, 131–135 (2002)
12. Gotthardt, U., Schweiger, U., Fahrenberg, J., Lauer, C.J., Holsboer, F., Heuser, I.: Cortisol, ACTH, and cardiovascular response to a cognitive challenge paradigm in aging and depression. Amer. J. of Physiol. 268, 865–R873 (1995)
13. Stratton, J.R., Levy, W.C., Caldwell, J.H., Jacobson, A., May, J., Matsuoka, D., Madden, K.: Effects of aging on cardiovascular responses to parasympathetic withdrawal. J. Am. Col. Card. 41, 2077–2083 (2003)
14. Schwartz, A.R., Gerin, W., Davidson, K.W., Pickering, T.G., Brosschot, J.F., Thayer, J.F., Christenfeld, N., Linden, W.: Toward a causal model of cardiovascular responses to stress and the development of cardiovascular disease. Psychosom. Med. 65, 22–35 (2003)
15. Kilander, L., Nyman, H., Boberg, M., Lithell, H.: The association between low diastolic blood pressure in middle age and cognitive function in old age. A population-based study Age. and Ageing 29, 243–248 (2000)
16. Kuo, H.K., Sorond, F., Iloputaife, I., Gagnon, M., Milberg, W., Lipsitz, L.A.: Effect of blood pressure on cognitive functions in elderly persons. The Journals Of Gerontology. Series A, Biological Sciences and Medical Sciences 59, 1191–1194 (2004)
17. Swan, G.E., DeCarli, C., Miller, B.L., Reed, T., Wolf, P.A., Jack, L.M., Carmelli, D.: Association of midlife blood pressure to late-life cognitive decline and brain morphology. Neurol 51, 986–993 (1998)

Monitoring Performance and Mental Workload in an Automated System

Indramani L. Singh, Anju L. Singh, and Proshanto K. Saha

Cognitive Science Laboratory, Department of Psychology
Banaras Hindu University, Varanasi – 221 005
ilsingh_bhu@rediffmail.com

Abstract. Human performance in computer-aided system has engrossed inevitably human issues in cognitive functioning. The present endeavor focuses on the associated influence of training, automation reliability on the monitoring performance and workload in multi-task ambience. MAT battery was utilized with engine-system monitoring, two dimensional tracking, and fuel resource management tasks were the concerned elements, in which only system engine-monitoring task was automated in the training as well as in the final test sessions. A 2 x 2 x 2 x 3, mixed factorial design was employed. Monitoring performance, false alarms, reaction time and root mean square error performance were recorded as dependent measures. Results revealed that automation-induced complacency might be the feature of multi-task condition where subjects detected automation failures under high static system reliability. Results further showed that mental workload significantly reduced from pre- to post-sessions.

Keywords: Automation, Complacency Workload, Monitoring Performance.

1 Introduction

Automation technology is a pervasive phenomenon, which has engrossed inevitably human issues of cognitive functioning. Automation envisages the thought of electronic replacement of human operator. Evidently, automation altered person's attitude towards the functioning of an automated machine, thereby placing more reliance and conviction on such systems [16]. Besides this another notion of automation has also been put forth as the execution of functions by a machine (preferably a computer) that was previously carried out by a human resource [11]. The performance in automation scenario depends on the interaction of people with the advanced technology. Moreover, the relationship between automation and mental workload of operator is an imperative consideration in respect to efficiency and safety in many modern human machine systems. For example, cockpit automation has made it possible to reduce flight times, increase fuel efficiency, navigate more effectively, and extend or improve the pilot's perceptual and cognitive capabilities [20, 21]. It is noteworthy that the benefits derived from automation use come after paying certain costs also [12], for instance, increased monitoring demands, unbalanced trust,

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 426–435, 2007.

cognitive overload, decision biases, skill degradation etc. Over trust of automation is sometimes referred to as complacency, which occurs when people trust the automation more than what is warranted and can result in very severe negative consequences, if the automation is less than fully reliable [12, 11]. Complacency can lead to a decreased monitoring of the system and a decreased likelihood of detecting system malfunctions. Moreover, introduction of automation in complex systems is embraced as reduction of the workload, and thereby reducing the human error. Instead, Woods [23] argued that automation merely changes how work is accomplished. Wiener [22] has even claimed that in some instances, the introduction of automation may increase the workload.

Bearing in mind such elements, researchers [12], put forth that any performance consequences for complacency were more likely to exist in a multi-task ambience. They examined the effect of levels of automation reliability i.e., constant and variable, on automated monitoring performance in various experimental conditions. Poorer monitoring efficiency was observed under high constant automation reliability compared to the variable. The results suggested that automation-induced complacency was more easily detectable in a multi-task environment when operators were engaged in performing numerous tasks. Considering the issue Singh et al. [18] presented automated task at the centre by spatial superimposition of monitoring task over the tracking, which could eliminate monitoring inefficiency (complacency). Results indicated that automation-induced complacency was not primarily influenced by the location, to be monitored for the automated task. Later, Singh, Molloy and Parasuraman [17], conducted another study and they showed that the centrally located monitoring task could not improve monitoring performance in automation mode, which suggested the robust nature of automation-induced complacency phenomenon.

Training is another important issue relevant to automation-induced complacency. Automation can place conflicting demand upon pilots, with which they may not be well trained to meet (e.g., passive monitoring versus active control) unless they have been specifically trained to cope with these demands. It has been suggested that inadequate training may lead to several automation-induced problems in the cockpit. For instance the negative effect of automation on monitoring performance may be related in part to a lack of ‘automation based’ skills. Recently, Sharma and Singh [14], examined the role of increased amount of manual training and automation reliability on a flight simulation task. Results indicated that, there was no benefit of extended manual training on automated complacency.

The rationale behind introducing automation in complex system is the reduction of workload and hence thereby reducing the human error propensity. Despite the logical and intuitive rationale that operators will choose automation under heavy workload, studies both of pilots and non-pilots performing laboratory [13], and aviation like tasks [3], revealed little, if any, tendency to choose automation more often at higher levels of task demand. It is possible that the influence of workload on automation use may emerge only when the workload is experienced for a sustained period of time. Another possibility is that more complex attributes of workload in real environments such as workload management and trade-offs need to be modeled in the laboratory in order to more fully comprehend the impact of workload on the use of automation. Thus, automated system can both reduce and increase mental workload. For instance, pragmatically it has been observed that glass cockpits in commercial aircraft have

relieved workload in areas such as reduced display clutter, and enhanced automated flight procedures [7]. However, the same cockpit systems can amplify workload by presenting operators with more options in their task and causing mode uncertainty [6]. Hence, these studies revealed that the automation might/or might not affect workload.

1.1 Present Study

The underlying principle behind the present endeavor focuses on the associated influence of extended automated training, automation reliability on the monitoring performance and workload in multi-task ambience. Some studies have suggested that automation ought to be designed with the objective to reduce operator's mental workload, while some other studies have indicated that automation does not necessarily reduce workload. In view of these contentious issues about the role of training, reliability and workload on the detection of automation failures, an effort has been made to examine the effects of extended automated training and system reliability on the relationship between monitoring automation failure and mental workload. Primarily, it was hypothesized that, increased automation training would reduce automation-induced complacency; secondly, automation induced complacency would be higher in constant system reliability condition than in variable system reliability along with increase over time periods in constant system reliability condition than in variable system reliability condition and, finally, automation would reduce mental workload.

2 Methods and Procedure

Participants: Eighty non-pilots with normal (20/20) or corrected to normal visual acuity, aged 19 to 25 years, volunteered in this study. Subjects were randomly assigned in each of the four experimental conditions. Each subject received 10-min manual practice on flight simulation task, besides 3-min demo.

2.1 Flight Simulation Task

A revised version of multi-attribute task battery (MATB), [2] was used in the present study. Multi-attribute task battery is a flight simulation package, comprising engine-system monitoring, two dimensional compensatory tracking, fuel resource management, communications, and scheduling tasks. The modified version of MATB allows each component task to be performed either manually or under automation mode. In the present study, only engine-system monitoring task was automated in the training as well as in the final test sessions. These three tasks were presented in separate windows on a 14" SVGA color monitor of a PC-486 computer.

2.2 NASA-Task Load Index Scale

NASA Task Load Index (NASA-TLX), [4] was administered before and after the final test session individually for the assessment of mental workload. The reliable index of overall was .83. In TLX, workload is defined as the 'cost incurred by human operators to achieve a specific level of performance.'

2.3 Design

A 2 (Training) x 2 (Automation reliability) x 2 (Sessions) x 3 (Blocks) mixed factorial design was employed with repeated measures on the last two factors. Training (30-min and 60-min) and automation reliability (constant and variable) were treated as between subject factors with sessions (two 30-min) and blocks (six blocks, each of 10-min.) as within-subjects factors. Automation reliability was defined as the percentage of correct detection of malfunctions by the automation routine in each 10-min block in the engine-system monitoring task. The rate of automation reliability was constant (87.5%) from block to block in constant system reliability and in the variable system reliability automation varied from high (87.5%) to low (56.25%) and low to high from block to block alternately. Each subject was administered six 10-min blocks in two successive 30-min sessions.

2.4 Procedure

Out of eighty non-pilots, 40 were given short- automation training and remaining 40 were given long training. Furthermore, out of 40 non-pilots in each training group, 20 were randomly assigned the constant system reliability and the remaining 20 subjects were assigned the variable system reliability.

In the automated test session, the subjects were informed that only engine-system monitoring task would be automated and they were instructed to pay attention to only on tracking and fuel management tasks. Subjects were also told that automation routine was less than 100% reliable and in case of automation failure they had to detect the malfunctions, if any, and to reset the system-monitoring task immediately, by pressing a designated function key within 10 seconds. The correct detection (hits rate), incorrect detection (false alarms) of malfunctions, monitoring reaction time (RT) and root mean square (RMS) error were recorded as dependent measures.

3 Results

Means and standard deviations for correct detection of automation failure on all components of MATB were calculated for six 10-min automated blocks. Mean values of the correct detection of monitoring task performance showed that the subjects detected slightly more malfunctions in the long-training-variable reliability condition (M = 59.15; SD = 33.98) than in its counterparts i.e. short-training-variable reliability condition (M = 58.98; SD = 33.98). Similarly, subjects detected high number of malfunctions in the long-training-constant reliability condition (M = 42.50; SD = 37.32) than in the short-training-constant reliability condition (M = 39.24; SD = 34.27).

The mean correct detection performance further demonstrated that the subjects detected slightly more malfunctions in the long training (M = 50.83, SD = 37.29) as compared to the short-training condition (M = 49.12, SD = 35.82), irrespective of system reliability. Furthermore, subjects' detection accuracy was higher in the variable system reliability condition (M = 59.07, SD = 34.03) than in the constant reliability condition (M = 40.87, SD = 35.66), irrespective of training. These mean performances further revealed that the mean correct detection performance decreased

from 3% to 15% after 30-min in the three experimental conditions except in the long-training-variable reliability condition. It is also evident from the mean performance that decrement in performance across blocks appeared from 23% to 40% in the constant reliability, whereas 4% to 5% deterioration emerged at some point of time among six blocks (see Fig. 1).

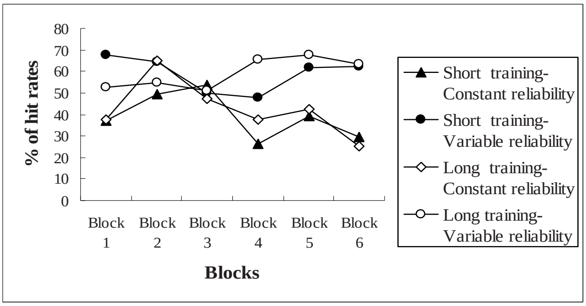


Fig. 1. Correct detection performance (hit rates) under four experimental groups during automated test sessions

Monitoring performance data were then computed for a 2 x 2 x 2 x 3 analysis of variance. The ANOVA showed that the main effect of training was not significant, $F_{(1, 76)} = .13$; ns, thereby suggesting that the amount of automation training given prior to the subjects under constant or variable automation reliability conditions has no impact on monitoring inefficiency (automation-induced complacency). Thus, the present finding does not support the first hypothesis that the high amount of automation training would reduce automation-induced complacency (see Fig. 2). Results further revealed that the main effect of system reliability was found highly significant $F_{(1, 76)} = 14.88$; $p < .01$. This result also indicated that the subjects had high reliance on automation while automation reliability was constant from block to block resulting in poor accuracy (more complacent). Contrarily, in variable reliability condition the reliance of subjects was varying from high to low and low to high, so subjects allocated more attention in detection of automation failures, resulting in better monitoring performance (less complacent). This finding supported second hypothesis, which maintained that the automation-induced complacency would be higher in the constant system reliability than it would be in the variable reliability.

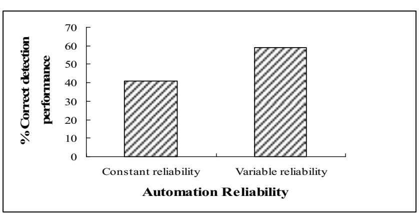


Fig. 2. Correct detection performance as a function of automation (system) reliability

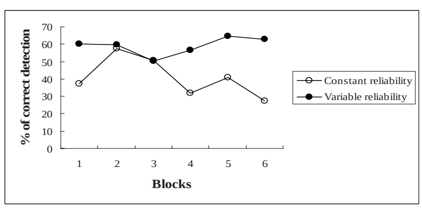


Fig. 3. Correct detection performance as a function of automation reliability across 10-min block

The interaction of system reliability and sessions, $F_{(1, 76)} = 9.39$; $p < .01$, was also found to be significant suggesting that types of system reliability i.e., constant and variable reliability impaired monitoring performance across sessions resulting in automation-induced complacency after a session of 30-min. The system reliability by session and by block interaction effect was also significant, $F_{(2, 152)} = 3.13$; $p < .05$ (see Fig. 3). The main effect of block was further found highly significant which propounded that automation-induced complacency appeared across time periods. This finding explained that automation-induced complacency increased across sessions and blocks, which supported the third hypothesis stating that automation-induced complacency would increase across time periods in constant reliability condition. The obtained findings on false alarms performance indicated that automation had no impact on performance however, the effect of system reliability suggested that the subjects committed more false alarms in the variable system reliability than in the constant reliability.

The ANOVA findings on reaction time performance showed that the main effect of training was significant, which suggested the benefit of the long training over short training. The system reliability and session interaction also reached at significance level, $F_{(1, 76)} = 4.65$; $p < .05$, which indicated that system reliability affected the speed of response in detecting correct automation failures in the flight simulation task across sessions.

The RMSE performance on tracking task suggested no benefit of training and system reliability. However, tracking performance showed an improvement across two sessions, $F_{(1, 76)} = 12.69$; $p< .01$. The ANOVA results on fuel performance indicated that the main effect of training was significant, $F_{(1, 76)} = 3.88$; $p< .05$. This finding suggested that the increased amount of training improved fuel performance. The results also indicated the benefits of either types of training over six 10-min blocks, $F_{(2, 152)} = 5.71$; $p< .01$. However, the main effect of system reliability was not significant, $F_{(1, 76)} = .003$; ns. This finding indicated no effect of system reliability on fuel resource management performance. The interaction effect of training and system reliability was also not significant, $F_{(1, 76)} = .58$; ns. This finding further showed that the length of training and system reliability could not enhance fuel performance.

3.1 Perceived Mental Workload

The NASA-TLX includes the process evaluating of relative importance of the six subscales by each subject to calculate a weighted mean. Miyake and Kumashiro [10], reported a high correlation ($r = 0.971$) between the weighted mean computed in the NASA-TLX and the simple arithmetic mean ratings of the scale. Their results suggested that the mean rating can be considered an appropriate subjective workload measure [1]. Similarly, Hendy, Hamilton and Landry [5], also suggested that weighting the ratings could not add to the sensitivity of the NASA-TLX. Thus, in the present study this evaluation process was omitted. A simple arithmetic mean was computed across subscales of the NASA-TLX and it was treated as a subjective mental workload score.

The NASA-task load index was administered to all the subjects at two times i.e. at pre- main task session and post-main task session. Each subject received two mental workload rating scores on six subscales of mental workload i.e. mental demand,

physical demand, temporal demand, effort, frustration and performance. The ANOVA results revealed that the main effects of mental demand ($F_{1, 76} = 16.59$; $p < .001$), temporal demand ($F_{1, 76} = 9.33$; $p < .01$), effort ($F_{1, 76} = 10.62$; $p < .01$), frustration ($F_{1, 76} = 16.59$; $p < .001$), performance ($F_{1, 76} = 73.29$; $p < .001$), and overall mental work load ($F_{1, 76} = 5.79$; $p <.01$) were significant except physical demand. These results suggested that mental demand (pre- M = 79.36; post M = 72.79), temporal demand (pre- M = 67.93; post M = 61.35), effort (pre- M = 74.08; post M = 66.91), frustration (pre- M = 32.00; post M = 21.76), and overall mental workload (pre- M = 60.99; post M = 58.75) were rated higher at pre-test than they were at their counterparts i.e. post test session, irrespective of the experimental conditions. However, subjects experienced almost equal physical demand at pre- and post-sessions (pre- M = 52.36; post M = 53.57). Moreover, rating of own performance workload (pre- M = 60.19; post M = 76.01) increased during test sessions. The performance workload is associated with the level of satisfaction the subjects felt about his/her performance in accomplishing the goal of the task. These findings supported the fourth hypothesis, which suggested that automation would reduce mental workload. To examine further the relationship between subscales of the NASA-TLX at pre- and post-test sessions, Pearson's product moment correlations were computed. The majority of correlations amongst the various measures of workload were highly significant with r-values ranging between 0.27 and 0.80. In sum, it is evident from the results that subjects experienced high mental workload on majority of the sub-scales of the NASA-TLX at the initial level i.e., pre-test, which got reduced across time periods while performing two 30-min flight simulation tasks (see Fig. 4).

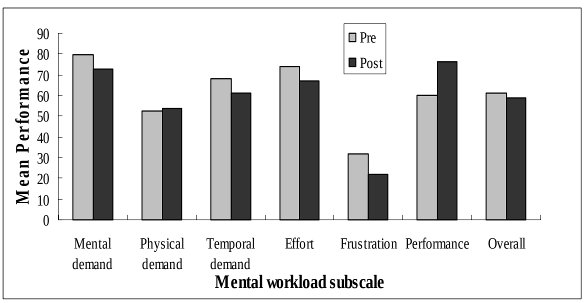


Fig. 4. Pre-and Post workload scores in NASA-TLX subscale

4 Discussion

Automation plays a critical role in situations when a small number of operators must control and supervise a very complex set of remote processes. Automation here is not optional; it is a necessity [15]. Complex machines tend to distance operators from the details of an operation. Over time, if the machines are reliable, operators will come to rely upon them, and may become less concerned with the details of the process. Though this has the desirable effect of moderating human operator workload, it also

has the undesirable effect of making the operator feel less involved in the task being performed. That's where complacency comes into play within the domain of cognitive downfalls. The workload overload consequences might affect the human performance from increases in task resource demand or stressful situations. These also may have some unforeseen errors like more selectivity of input, more important sources of information given more weight, decrease in accuracy, decreasing use of strategies that involve heavy mental computation and locking into a single strategy. Most critical being the operators continuing awareness of the objective importance of all tasks that may compete for attention, for instant those of lesser importance will be shed first. Therefore the antidote may include to redesign the task by assigning some loaded tasks to automated mode and include a display design such that, information for the most tasks are available, interpretable and salient.

Several reports have discussed the dangers of automation-induced complacency. The general experience in aviation has been that advanced automated devices often do reduce workload, but usually at flight phases where workload is already low, such as cruise whereas some automation actually increases workload at critical phases, for instance at the time of take off and landing. Thus automation merely shifts the pattern of workload between work phases. However, little empirical research has been produced to substantiate its harmful effects on performance as well as some other cognitive factors that could be the root cause for automation-induced complacency. The present endeavor looks into revalidation of earlier findings of automation-induced complacency [12, 19], and also an examined the intriguing relationship among extended training reliability and mental workload. The current experimental results suggest that automation-induced complacency might be the feature of multi-task condition where subjects detected automation failures under high static system reliability condition as compared to variable system reliability condition. Imperatively, this effect of automation-induced complacency further enhances after half an hour of task period.

Considering further, the present experimental conditions also looked into the discrete relationship between subjective mental workload and automation-induced complacency. The obtained pre and post mental workload revealed that mental demand, temporal demand, effort, frustration and overall workload significantly reduced from pre to post sessions. The performance-rating workload of subjects, further significantly enhanced over time, whereas physical demand was stable during test sessions. The compilation of results suggested that high system reliability would reduce workload, thereby resulting in automation-induced complacency. And hence this finding corroborated the findings of Metzger and Parasuraman [9], who recently reported that reliable automation reduces mental workload.

5 Conclusion

Automation is rapidly being incorporated in modern flight decks thereby accounting for augmented system reliability and decreasing operational errors. However, despite it's potential for improving overall system (or aircrafts) performance, automation is not always considered good. It has been always associated with unique and unanticipated problems. Automation has enhanced the importance of the need to

comprehend and control the factors that influence human monitoring behavior. Errors are least likely to show their effect or influence the results, when workload is moderate and does not alter suddenly or unpredictably [8]. Thus, this study argues for the goal to emphasize consequences of automation. Maintaining appropriate levels of workload during automated operating conditions is one of the key issues in the design of nuclear power plants and other process control environments. Therefore, over and under load continue to be a critical cognitive and human factors issue. Additional studies with different monitoring tasks, ingenious methods for adaptive control and under different scenarios of multiple task performance need to be demeanor to test effectiveness of sophisticated approaches.

References

1. Byers, J.C., Bittner, A.C., Hill, S.G.: Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In: Mital, A. (ed.) Advances in industrial ergonomics and safety I, Taylor & Francis, London (1989)
2. Comstock, J.R., Arnegard, R.J.: The multi-attribute task battery for human operator workload and strategic behavior research. (NASA Tech. Memorandum 104174). NASA Langley. Research Centre, Hampton, VA (1992)
3. Harris, W., Hancock, P.A., Arthur, E.: The effect of task load projection on automation use, performance, and workload. In: Proceedings of the 7th International Symposium on Aviation Psychology, Columbus, OH (1993)
4. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of experimental and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human mental workload, pp. 139–183. Human mental workload, North Holland, Amsterdam (1988)
5. Hendy, K.C., Hamilton, K.M., Landry, L.N.: Measuring subjective workload: When is one scale better than many? Human Factors 35, 579–601 (1993)
6. Hilburn, B.: Dynamic decision aiding: the impact of adaptive automation on mental workload. In: Harris, D. (ed.) Aldershot: Ashgate, pp. 193–200. Ashgate, Aldershot (1997)
7. Hughes, D., Dornheim, M.A.: Accident’s direct focus on cockpit automation. Aviation Week. and Space Technology 30, 52–54 (1995)
8. Kantowitz, B.H., Casper, P.A.: Human workload in aviation. In: Wiener, E., Nagel, D. (eds.) Human factors in aviation, pp. 157–185. Academic Press, New York (1988)
9. Metzger, U., Parasuraman, R.: Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. Human Factors 47(1), 35–49 (2005)
10. Miyake, S., Kumashiro, M.: Subjective mental workload assessment technique-an introduction to NASA-TLX and SWAT and a proposal of simple scoring methods. Japanese Journal of Ergonomics 29, 399–408 (1993)
11. Parasuraman, R., Riley, V.: Humans and automation. Use, misuse, disuse, abuse. Human Factors 39, 230–253 (1997)
12. Parasuraman, R., Molloy, R., Singh, I.L.: Performance consequences of automation-induced complacency. International Journal of Aviation Psychology 3(1), 1–23 (1993)
13. Riley, V.: A theory of operator reliance on automation. In: Mouloua, M., Parasuraman, R. (eds.) Human performance in automated systems: Current research and trends. LEA, Hillsdale, NJ, pp. 8–14 (1994)

14. Sharma, H.O., Singh, I.L.: Extended training, automation reliability and monitoring task performance. Psychological Studies, 49 (4) (2004)
15. Sheridan, T.B. (ed.): Humans and automation. System design and research issues. Wiley, New York (2002)
16. Singh, I.L., Molloy, R., Parasuraman, R.: Automation-induced complacency. Development of the complacency potential rating scale. International Journal of Aviation Psychology 3, 111–122 (1993)
17. Singh, I.L., Molloy, R., Parasuraman, R.: Automation-induced monitoring inefficiency: Role of display location. International Journal of Human Computer Human Computer Studies 46, 17–46 (1997)
18. Singh, I.L., Molloy, R., Parasuraman, R., Westerman, S.: Does location affect automation-induced complacency? In: Paper presented in the First Conference on Automation and Human Performance, Washington, DC, USA (1994)
19. Singh, I.L., Parasuraman, R.: Human performance in automated systems. Indian Psychological Abstract and Reviews 8(2), 235–276 (2001)
20. Singh, I.L., Sharma, H.O., Singh, A.L.: Effect of training on workload in flight simulation task performance. Journal of the Indian Academy of Applied Psychology 31(1-2), 81–90 (2005)
21. Wiener, E.L.: Cockpit automation. In: Wiener, E.L., Nagel, D.C. (eds.) Human factors in aviation. pp. 433–461. Academic, San Diego (1988)
22. Wiener, E.L.: Human factors of advanced technology ("glass cockpit") transport aircraft (Report No.177528). Ames Research Centre, Moffett Field, CA (1989)
23. Woods, D.D.: Automation. Apparent simplicity, real complexity. In: Mouloua, M., Parasuraman, R. (eds.) Human performance in automated systems: Current research and trends, pp. 1–7. Lawrence Erlbaum Associates, Hillsdale, NJ (1994)

Context-Aware Notification for Mobile Police Officers

Jan Willem Streefkerk, Myra van Esch-Bussemakers, and Mark Neerincx

TNO Human Factors, Kampweg 5,
3769 DE Soesterberg, the Netherlands
{j.w.streefkerk, myra.vanesch, mark.neerincx}@tno.nl

Abstract. To minimize unwanted interruption and information overload during surveillance, mobile police officers need to be supported by a mobile, context-aware notification system. This system adapts message presentation to message priority and context of use. A prototype is designed and evaluated in a simulated surveillance task, requiring users to attend to videos while handling messages on a mobile device. Adaptive notification led to better performance and less intrusive messages than non-adaptive notification, especially in high workload situations. Subjective judgments showed a positive user experience with the adaptive notification system. These empirical findings are used to improve the design of mobile notification support systems for police officers.

Keywords: Mobile device, notification, context-aware computing, Wizard-of-Oz, police.

1 Introduction

Police officers on surveillance need to be aware of task-relevant information. For these mobile users, both the interaction possibilities with devices and the momentary user needs for information continuously change over time and place [1]. Location-based and context-aware services on mobile devices offer new possibilities for supporting the police officer's task such as providing up-to-date information in the relevant context (i.e. at a specific location). Both theory and police practice show a clear need for interfaces that attract and guide the attention of individual officers to relevant, high priority information in a mobile setting [7].

Designing such interfaces should take into account divided attention during surveillance. An intrinsic part of police officers' task is dividing their attention between detecting incidents in their environment and attending to incoming messages on a mobile device. Two potential risks to effective surveillance are unwanted distraction by low priority messages and information overload in high workload situations. These situations are caused by monitoring and integrating complex information from the environment, while simultaneously performing tasks on a mobile device [6]. Imagine a police officer on surveillance handling an argument between two individuals. During this intervention, his mobile device interrupts him with a message. He has to attend to and read the message to perceive its urgency and content. Focusing his attention on his device distracts him from the argument, which could have serious consequences. When he can perceive the urgency of a message by

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 436–445, 2007.

its presentation, i.e. by auditory signals, unwanted distraction will be minimized. Similarly, when the message content on the mobile device is presented concisely instead of elaborately, information presentation is optimized for the situation. Thus, message presentation on a mobile device must be matched to message priority and user workload.

1.1 Adaptive Notification Systems

Context-aware notification systems can provide this match by adapting information presentation based on a model of relevant use context factors. For example, the information provided by a context-aware mobile device is tailored to the user's location or resources [2]. No such support system has yet been developed for mobile police officers [7]. In earlier research, context-aware support for firefighters was designed to generate user alerts on a (Personal Digital Assistant) PDA based on sensor information, such as location, temperature, toxicity etc. Results of a formative user study showed that firefighters found the application useful [3]. In the police domain, a context-aware communication system was designed on a mobile device to support handling criminal incidents. Based on incident characteristics, contact details of colleagues who could assist with the incident were presented together with the message [9]. Results from a preliminary user study suggested that police officers found this application useful for efficient handling of incidents. Despite these promising results, there is still a lack of experimental evaluation of context-aware applications in task-relevant settings [10].

Our approach is designing a context-aware notification system on a mobile device that presents messages in the appropriate notification style, given message priority and user workload. The notification style determines message presentation, i.e. the auditory and visual signals and text in the user interface. For example, in a high workload situation, a medium priority message will be presented with another type of auditory signal than a high priority message. A semi-functional prototype of this context-aware notification system is implemented on a mobile device and evaluated in a simulated surveillance task. In this paper, first the design and implementation of the context-aware notification system is described. Then, the evaluation results are discussed and implications for design of notification systems are presented.

2 Design of Notification System

From the discussion above, we concluded that a context-aware notification system that adapts the notification style to context factors can support mobile police officers. In this section, we specify the notification styles, match these to the context factors of message priority and user workload and describe the implementation of a prototype.

2.1 Designing Notification Styles

The general goal of a notification system is to successfully prompt the user to attend to another task. Successful notification systems should facilitate 1) successful interruption, 2) comprehension of the notification and 3) quick user reaction [5].

Users should be able to perceive a notification, understand the message and act on the notification quickly and efficiently.

The notification styles were determined by two important aspects of notification: *salience* and *information density* [4]. We matched message priority to salience, because highly salient notifications are perceived more easily. Thus, high priority messages were presented highly salient (with visual flashing effects and loud auditory signals) and medium and low priority messages less salient. This was expected to benefit interruption and comprehension of notifications.

Information density is the amount of information conveyed by the notification, for example the amount of text in a display or the richness of an auditory signal. We matched user workload to information density, because the amount of information that has to be processed is an important factor in determining the speed of the response. Thus, more condensed information was presented in high workload situations. This was expected to benefit response time to notifications.

Concluding from these design considerations, the adaptive behavior of our notification system was dictated by two rules:

1. *If message priority is high, then use highly salient warning signals.*
2. *If user workload is high, then present more condensed information.*

2.2 Implementation of Prototype

A semi-functional prototype of this adaptive notification system was implemented on a mobile device (PDA) for evaluation. High priority messages were presented with red visual flashing bars and icons and a sharp sound (🔊) (see Table 1). Medium priority messages were presented with a soft sound (☼) and low priority messages without sound. In low workload situations the full message text was presented at once, but in high workload situations first a summary of the message was presented. To avoid interruption while users were interacting with the device (e.g. reading a message), in these situations an icon signaled a new message (right column in Table 1). An exception to the rules was made for a high priority message in high workload situations. Because we assumed that all relevant information is necessary to take action on high priority messages, the entire message was presented instead of a summary. The evaluation compared the context-aware prototype to a non-adaptive prototype presenting the full message text with a sharp sound for all messages.

3 Evaluation of Context-Aware Notification

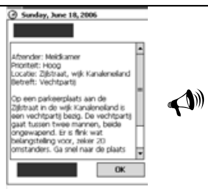
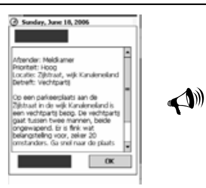
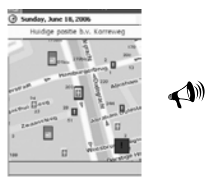
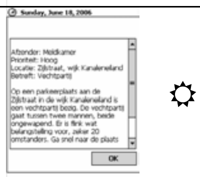
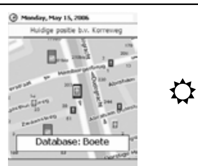
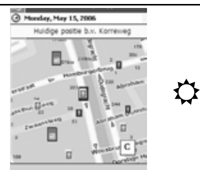
This lab study evaluated the prototype in a Wizard-of-Oz setup that simulated the adaptive system behavior [8]. This setup allowed varied system behavior between conditions and accurate measurements without the need to implement a fully functional system. By comparing context-aware notification to a non-adaptive prototype (i.e. uniform notification style), participants experienced the dynamics of this system. A representative participant group was used in the evaluation, because adaptive notification is expected to address general instead of police task-specific abilities.

A simulated, mobile surveillance task was created in which participants had to perceive, read and remember messages on a mobile device, while attending to videos. We expected that adaptive notification lead to better task performance during high workload situations, less interruption by low priority messages and higher preference compared to non-adaptive notification. Furthermore, we examined whether adaptive notification was recognized easily by users and whether the underlying rules for adaptive notification were appropriate.

3.1 Method

Participants. 20 participants (10 male and 10 female) between 22 and 45 years old (M=29.1; SD=6.8) completed all tasks in the experiment. All used a PC and the Internet on a daily basis, 17 participants used a cellular phone daily and most of them (N=19) had never before used a mobile device. None had working experience in the police domain.

Table 1. Matching notification styles to message priority (vertical) and context of use (horizontal)

	Attention on environment		**Attention on device**
	Low Workload	**High Workload**	
High priority			
Med priority			
Low priority	 —	 —	 —

Experimental Design. A within subjects, repeated measures design was used with an adaptive notification condition (notification style adapted to context of use and message priority) and a non-adaptive condition (uniform notification style). In both conditions, similar surveillance scenarios were used with similar messages and videos. The order of conditions and scenarios was counterbalanced.

Manipulation. In the adaptive notification condition, notification styles of messages were adapted to the context of use and the message priority. Context of use was characterized as low workload or high workload (determined by the location of the participant) or interacting with the device (taking notes or reading a message on the PDA). Workload (WL) was induced by videos shown in two "districts" represented by two adjacent rooms. Low workload videos (district 1) showed only bicyclists riding through a shopping street, whereas high workload videos (district 2) showed fights, arguments and attacks on police officers (see Figure 1). These videos required focused attention from the participant to accurately remember perpetrator and situation characteristics. When participants physically moved from one district to the other, prompted by the messages, this constituted a switch in the context of use.

All messages were relevant, however the priority of the messages differed between high, medium or low priority. High priority messages, such as a colleague in need of backup, were considered urgent and needed immediate attention. Medium priority messages (e.g. about an outstanding warrant) did not require immediate attention. Low priority messages, such as questions from colleagues on past events, were not classified as urgent.

Tasks. The mobile surveillance task consisted of moving between both districts, watching the videos and reporting targets from the videos. In district 1 (low workload) the number of bicyclists that rode into a shopping street had to be counted. In the other district (high workload), participants took notes on the PDA about the perpetrator and situation characteristics in the videos. In total 18 targets had to be identified from these videos. One scenario consisted of alternating three low workload videos and three high workload videos and took about 25 minutes to complete.

Fig. 1. Video stills from the low workload (left) and high workload district (middle and right)

The second task was handling the messages. A total of nine (three low, four medium and two high priority) messages were sent to the participants, spread evenly over the three different contexts. High priority messages had to be read as fast as possible and prompted participants to move from district 1 to district 2. Medium and low priority messages could be read when possible. To read the whole message on the PDA, scrolling was required.

Measures. In this experiment, we measured efficiency and effectiveness of task performance, mental effort, message intrusiveness and preferences for adaptive notification. Prior to the experiment, a questionnaire was filled out concerning gender, age, education, computer experience and use of mobile devices. During each scenario, task performance was measured as the *number of targets* reported from both the high and low workload videos. In addition, *message handling time* was measured in seconds. After each scenario, *mental effort* was measured with the Rating Scale Mental Effort [10] and *number of messages* recalled correctly with six multiple choice questions.

After each message, participants were asked to rate the *message intrusiveness* using a small, 7-point rating scale on the PDA. After each scenario, subjective judgments are measured with a 5-point rating scale on the ease of the surveillance task, difficulty in directing attention, interruption and irritation experienced by the messages. In addition, *recognition of adaptation* was measured with open questions after the adaptive condition asking participants which differences they noticed between the messages and contexts. Finally, after the experiment participants stated *preferences* by comparing both conditions and offered improvements to the prototype.

Apparatus. The adaptive notification system was simulated using a HP 5550 PDA. A Wireless LAN connected the PDA to a message server, used by the test leader to send the messages. The server kept an event log with time codes of all events on the PDA (messages received, messages opened, etc.). Videos were projected full-screen on the wall of the two rooms using two beamers controlled by a desktop PC. The rooms, PDA screen and the video images were videotaped as back-up.

Procedure. At the start of the experiment, participants were instructed they had to perform two surveillance scenarios by watching videos and handling messages on a mobile device. In one session, the notification style of the messages on the PDA would be adapted to their context of use, either "relaxed" (district 1, low workload), "busy" (district 2, high workload) or "working with the device". Also, the style would be adapted to the message priority. In the other session, all notification styles would be uniform. Participants could decide for themselves when to read a message.

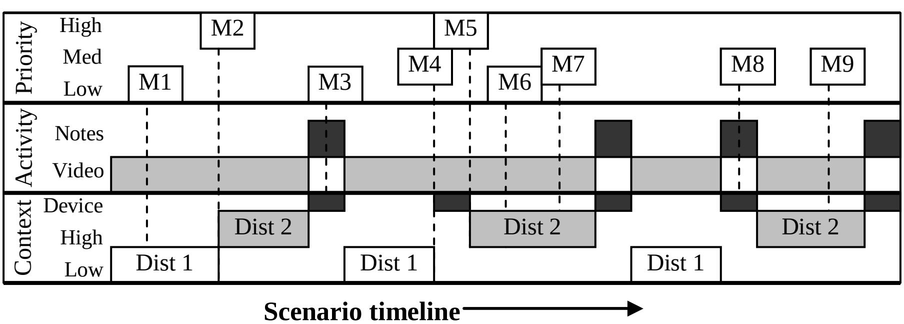


Fig. 2. Timeline of both scenarios with the presentation moments of the messages

After training with the PDA, messages and notification styles, participants performed both sessions with a short break in between. The timeline of the session is presented in Figure 2. After each session, they filled out the multiple choice questionnaire, recognition questionnaire and subjective rating scales. After both sessions, they were asked which session they preferred (session 1 or 2). After this, they were debriefed and paid for their participation.

Statistical Analyses. Repeated measures ANOVA with factors condition (2 levels) and context (3 levels) or message priority (3 levels) and Bonferroni post-hoc comparisons analyses were performed on message handling time and message intrusiveness. Number of targets and messages and mental effort are analyzed with separate t-tests for dependent samples. Finally, the questionnaires and rating scales were analyzed with non-parametric tests (Wilcoxon Matched Pairs). In the adaptive condition, message M6 (low priority, high workload) was only perceived and read by 4 participants and therefore excluded from the analysis.

3.2 Results

Number of Targets and Messages. Results indicate that there was no difference in the number of targets in the low workload context ($t(19)=1.04$; $p=.31$). However, significantly more targets were recalled in the high workload context ($t(19)=2.15$; $p<.05$). There was no difference in number of messages recalled correctly ($t(19)=0.21$; $p=.84$). Thus, task performance on the surveillance task profited from the adaptive system in high workload situations.

Message Handling Time. Message handling time was significantly different between conditions ($F(1,19)=35.1$; $p=.000$) and between contexts of use ($F(2,38)=18.4$; $p=.000$) (see Figure 3). The significant interaction effect ($F(2,38)=5.37$; $p<.01$) showed that message handling time was varied across conditions and contexts. Post-hoc comparison of MHT in low and high workload situations showed significant differences in the adaptive condition ($p=.000$) but not in the control condition ($p>.05$). Thus, message handling time was more varied in the adaptive condition.

Message handling time was also significantly different for different message priorities ($F(2,38)=29.3$; $p=.000$) (see Figure 3). Again, post-hoc comparison of MHT for low, medium and high priority messages showed significant differences in the adaptive condition (all $p=.000$) but also in the control condition ($p<.05$ and $p<.005$). However, differences were greater in the adaptive condition.

Table 2. Averages for the performance measures message handling time (*MHT)*, message intrusiveness *(MI)*, mental effort *(ME)*, number of messages *(#M)*, and number of targets *(#T*)

<table>
<tr><th>Condition</th><th>Context</th><th>MHT</th><th>MI</th><th>ME</th><th>#M</th><th>#T low</th><th>#T high</th><th>Priority</th><th>MHT</th></tr>
<tr><td rowspan="3">Adaptive</td><td>Low WL</td><td>27.3</td><td>1.9</td><td rowspan="3">46.7</td><td rowspan="3">3.3</td><td rowspan="3">12.4</td><td rowspan="3">13.1</td><td>Low</td><td>50.0</td></tr>
<tr><td>High WL</td><td>64.3</td><td>4.0</td><td>Med</td><td>52.3</td></tr>
<tr><td>Device</td><td>36.0</td><td>2.0</td><td>High</td><td>16.5</td></tr>
<tr><td rowspan="3">Control</td><td>Low WL</td><td>23.8</td><td>2.0</td><td rowspan="3">50.9</td><td rowspan="3">3.2</td><td rowspan="3">11.4</td><td rowspan="3">11.6</td><td>Low</td><td>14.9</td></tr>
<tr><td>High WL</td><td>39.3</td><td>5.1</td><td>Med</td><td>31.2</td></tr>
<tr><td>Device</td><td>14.8</td><td>3.3</td><td>High</td><td>18.0</td></tr>
</table>

Message Intrusiveness and Mental Effort. The messages were experienced as more intrusive in the control condition ($F(1,19)=17.6$; $p=.000$). In addition, post-hoc comparison showed that messages in the adaptive condition were experienced as less intrusive, especially when the device was in use ($p=.000$) and in high workload situations ($p=.000$). Thus, with the adaptive system, the intrusiveness of messages is lower. No significant differences were found for mental effort between conditions. However, a trend towards lower mental effort in the adaptive condition could be observed ($t(19)=1.93$; $p=.06$).

Subjective Judgments and Preferences. Results indicated that participants did not find the surveillance easier in one of both conditions ($Z(20)=1.48$; $p>.05$). However, in the adaptive condition they found it significantly easier to direct their attention between the messages and the videos ($Z(20)=2.56$; $p<.05$), were less interrupted and irritated by the messages ($Z(20)=2.28$; $p<.05$ and $Z(20)=2.82$; $p<.01$ respectively). Their preference was almost unanimously for the adaptive condition (90%) because they could distinguish message priority (50%) and the messages were less interruptive (25%). These results show high preference and positive judgments for adaptive notification.

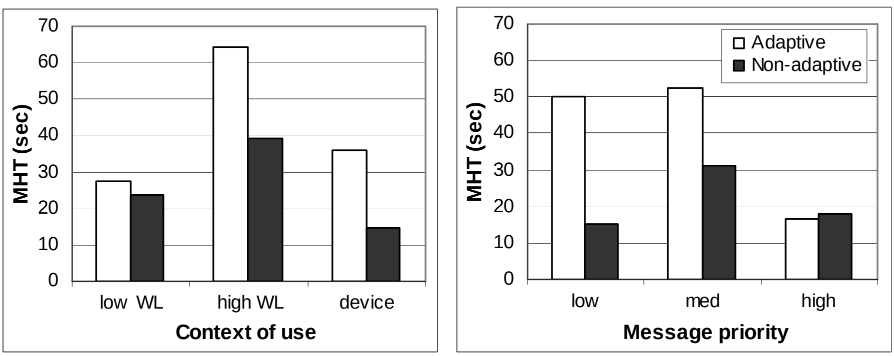


Fig. 3. Message handling time (*MHT*) for context of use (left) and message priority (right)

Recognition of Adaptation. In total 22 remarks were made about the differences in message presentation between the districts. Of these, only 23% identified the correct differences (district 1: whole message, district 2: summary). 36% of the answers were wrong or did not state any differences. The rest (41%) of the remarks did not concern message presentation. When asked about the differences in message presentation between the conditions, 25 correct remarks were made. Of these, 20% stated all differences, 24% were only about differences in priority, 24% only about visual aspects and 20% only about the auditory signals.

3.3 Discussion and Conclusions

This study evaluated adaptive notification compared to non-adaptive notification in a simulated surveillance task. Results showed that with the context-aware notification system, task performance was better than with the non-adaptive system, especially in

high workload situations. In these situations, more targets were reported from the high workload videos and message handling times were longer, showing participants postponed reading messages. Furthermore, with the adaptive notification the intrusiveness of messages was rated lower than the intrusiveness of the same messages presented with a non-adaptive system. When using the adaptive system, users could direct their attention better and felt less interrupted by the messages than with the non-adaptive system. The preference was almost unanimously in favor of the adaptive notification system. These empirical findings support the anticipated usefulness of context-aware notification systems and positive subjective judgments found in earlier research in other domains [3].

Overall, the notification system allowed users to quickly perceive, comprehend and act on messages, suggesting the notification system was appropriately designed for this task. The variance in message handling time (short for high priority, longer for low and medium priority) showed participants could recognize message priority correctly from the notification style. The correct answers on the recognition questionnaire support this conclusion. Participants did remark that low priority messages should also be announced by an auditory signal. In addition, the current prototype uses the visual modality heavily, keeping visual attention away from the environment. Instead, the auditory modality could be used to provide more information-rich signals, for example with synthesized speech.

Many participants did not notice and recognize the differences in message adaptation between contexts. It was not clear to participants that the appearance of a summary or icon depended on their location or activity. However, participants found the summary before a message useful and task performance was better than with full-text messages. This shows that the matching between user workload and information density was intuitive and appropriate. The adaptive system behavior supported the users' task flow, even though users may not actively notice it. Thus, the adaptation of information density is useful for notification systems.

The Wizard-of-Oz setup was successful in creating the illusion of a functioning adaptive system. In addition, using videos and messages successfully created a divided attention situation, characteristic of a surveillance task. However, participants only needed to remember the content of the messages and the videos and were not required to participate actively in the situations. A more elaborate and immersive task environment, such as a game-based environment, can also take into account the task flow, situation awareness and the dynamics of the user experience.

In conclusion, this study provided empirical data on the impact of a context-aware notification system in a task-relevant setting. When using this system, task performance was better in high workload situations and users felt less interrupted than with a non-adaptive system. We continue this ongoing research by designing and testing these systems in game-based environments and in field studies with end-users. By relating these results back to the design process, usable and innovative support concepts for police officers are realized.

Acknowledgements. The MultimediaN project is sponsored by the Dutch Ministry for Economic Affairs. We thank Bert Bierman for programming the prototype and Natasha Weitenberg and Cirquest for the video material.

References

1. Baber, C., Haniff, D., Sharples, M., Boardman, M., Price, A.: A requirements analysis of personal mobile computers for police officers. In: Harris, D. (ed.) Engineering Psychology and Cognitive Ergonomics, Volume Six: Industrial Ergonomics, HCI, and Applied Cognitive Psychology, pp. 21–28. Aldershot, Ashgate (2001)
2. Jameson, A.: Usability Issues and Methods for Mobile Multimodal Systems. In: Proceedings of the ISCA Tutorial and Research Workshop on Multi-Modal Dialogue in Mobile Environments (summary of a keynote address), Kloster Irsee, Germany (2002)
3. Jiang, X., Chen, N.Y., Hong, J.I., Wang, K., Takayama, L., Landay, J.A., SIREN: Context-aware Computing for Firefighting. In: Proceedings of Pervasive 2004. Vienna (2004)
4. Kern, N., Schiele, B.: Context–aware notification for wearable computing. In: Proc. 7th International Symposium on Wearable Computers, pp. 223–230. New York, IEEE (2003)
5. McCrickard, D.S., Catrambone, R., Chewar, C., Stasko, J.: Establishing tradeoffs that leverage attention for utility: empirically evaluating information display in notification systems. Int. J. Human-Computer Studies 58, 547–582 (2003)
6. Nagata, S.F.: User Assistance for Multitasking with Interruptions on a Mobile Device. Unpublished thesis 7, (2006)
7. Streefkerk, J.W., van Esch-Bussemakers, M.P., Neerincx, M.A.: Designing Personal Attentive User Interfaces in the Mobile Public Safety Domain. Computers in Human Behavior 22, 749–770 (2006)
8. Streefkerk, J.W., van Esch-Bussemakers, M.P., Neerincx, M.A.: Evaluating context-aware mobile user interfaces. In: Lumsden, J. (ed.), Handbook of research on user interface design and evaluation for mobile technology. Idea Group (in press)
9. Van Eijk, R., De Koning, N., Steen, M., Reitsma, E.: Developing a We-Centric Service for Mobile Police Officers to Support Spontaneous Communication. Freeband, Enschede (2006) (available online at https://doc.freeband.nl/dscgi/ds.py/Get/File-67631)
10. Weibelzahl, S.: Problems and pitfalls in the evaluation of adaptive systems. In: S. Chen, G. Magoulas (eds.): Adaptable and Adaptive Hypermedia Systems. IRM Press, pp. 285-299 (2005)
11. Zijlstra, F.R.H.: Efficiency in work behavior. In: A design approach for modern tools (PhD thesis, Delft University of Technology), Delft University Press, the Netherlands (1993)

A Study on the Vertical Navigation of High Rise Buildings

Xianghong Sun[1], Tom Plocher[2], and Weina Qu[1]

[1] State Key Laboratory of Brain and Cognitive Science, Inst. of Psychology,
Chinese Academy of Science,
Beijing 100101, China
{sunxh, quwn}@psych.ac.cn
[2] Honeywell ACS Labs,
Minneapolis MN 55418, USA
tom.plocher@honeywell.com

Abstract. Scroll bar and stab touch screen controls were designed and compared to a soft keyboard to help firefighters perform vertical navigation tasks for high-rise buildings depicted on a graphical display. 18 male subjects were asked to accomplish three experimental tasks: 1) input floor number and navigate to the floor, 2) switch the current floor to another one that was two floors higher/lower, and 3) check around the floors in the high rise to find the one on fire. Task completion time and number of screen operations were recorded. Results showed that, keyboard method of floor selection was the fastest, and scroll bar the slowest. In Task 1 and Task 2, the least number of screen operations were shown with soft keyboard. But in Task 3, keyboard was slowest. The other two control methods were not sensitive to tasks. Design implications for scroll bar and stab controls are discussed.

Keywords: touch screen, vertical navigation, scroll bar, graphical displays.

1 Introduction

In order to detect the original fire and the current fire spread in a high rise building as quickly and accurately as possible, intelligent fire information display systems should show firefighters detailed information floor by floor, such as the floorplan, the location of a fire, the current position of firefighters, and the locations of fire hydrants, water pipes, stairs, and dangerous materials. An example of such a display is shown in Figure 1. The fire information display must allow firefighters to switch between floors very easily to view the floor plan and associated information objects. What is an intuitive way for firefighters to input the desired floor number or navigate vertically through the floors of the building searching for the source of the fire? The situation is analogous to that of people using an elevator. But, unlike an elevator control panel, a computer touch screen does not have unlimited space for buttons. For buildings higher than ten floors, there simply is not enough screen space to assign a single button to each individual floor.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 446–454, 2007.

Vertical navigation techniques that are effective, but much more efficient in their use of screen space are needed[1,5,6]. In this study, we designed several kinds of vertical navigation controls for a touch screen display, and determined which one was the best for fire fighters to use for gathering information from the graphical floor plans.

From existing computer interface designs, we have found several possible ways to support the fire fighter's task of switching between floors and floor plans[2,3].

First, the *scroll bar* has been widely used for moving between different parts of a document. For our vertical scroll bar control, we assumed that the location of the slide represents the floor level. The higher the slide position is on the scroll bar, the higher the floor level. It bears a quite strong analogy to a real high-rise building that we assumed should be reflected in ease of use.

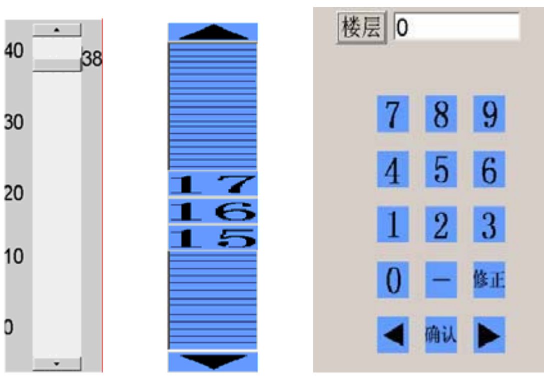
Fig. 1. Typical firefighter information display (with experimental scrollbar control)

A second approach is the *stab control*. Typically, the stab control is used in a sub-menu to list all the functions when the functions and operations are too numerous to display in one page and when some of them can be grouped together. Typically, the stab control is presented in a horizontal way. For our application it was positioned vertically. One stab was assigned to each floor level. The selected floor plan was shown in full size, and the others half hidden and overlapped behind the selected floor plan. The user switches floor levels by clicking the corresponding stab on the screen.

Keyboards, our third approach, are used to accurately input or set a value, such as setting the font size at 11.5 points. With mobile phones, access control systems, etc., it is easy to find keyboards in daily life. Even people who lack computer experience know how to use a keyboard to input digits. If a fire fighter wanted to go to a specific floor, such as the 14th floor, he or she would simply type in "14", then press "enter". But one problem with keyboards is that it is difficult to change the current number to another one even if there is only one digit difference between the two numbers. One would question how suitable it is for a task that frequently involves rapidly changing the number up and down one digit at a time. We would predict that it is not very well suited to the fire fighter display task of switching floors.

Based on interviews with Chinese firefighters, it was found that they have three main tasks that require vertical navigation through the building floor plans: 1) to check floor by floor and find which floors are involved in the fire, 2) to switch between floors

and find the stairways, and 3) to go to a specific floor for a detailed description of key objects. The three vertical navigation designs are shown in Figure 2. The left one is the scroll bar that includes a movable slide and two up/down arrow buttons. The slide could be dragged to any specific floor level and the current floor number shown beside the slide. The up/down arrow buttons could make the slide one floor higher or lower. The middle one is the stab display that looks like a set of cards shown vertically on screen. The current floor and two adjoined floors are shown on cards. The other cards were overlaid behind the current floor card. There was also a pair of up/down arrow buttons at the two ends of the stab to make the current floor card go up/down by one floor. To the right is shown the soft keyboard display that included 10 digit buttons and two left/right arrow buttons. People can press the digits on the screen to get to a specific floor, and press the left/right arrow to go up/down by one floor.



(a) scroll bar (b) stab control (c) soft keyboard on screen

Fig. 2. Three vertical navigation control designs

Comparing the three kinds of vertical navigation methods, the following questions need to be answered. 1) Which method is the fastest for helping firefighters select a specific floor? 2) Which method is the fastest for switching between floors? 3) Which method is the easiest for moving up/down by just one floor? We hypothesized that since the scroll bar is easy to learn and has a strong analogy to a high-rise building, it would be the best method for the vertical navigation.

2 Method

2.1 Experimental Design

The independent variable consisted of three kinds of vertical navigation methods: keyboard, vertical scroll bar, and stab control. The dependent variables were task completion time and the time to complete screen operations for each task, including clicking a button, and dragging a finger on screen.

2.2 Experimental Tasks

Corresponding to the three questions mentioned above, there were three experimental tasks:

Task 1. Move to a specific floor level by inputting the floor number as directed.

Task 2. Switch to another floor level that is two floors higher or lower than current position. This task was designed for firefighters to check detailed information at adjacent floors. Different from the Task 1, this task displayed the current floor number on the screen, and asked subjects to change the current floor to another one. For the soft keyboard display, the subjects had to delete the current number, and then input another one. Another way to use the keyboard for this was to click the left/right arrow button to make the floor number increase/decrease by one floor. For the scroll bar, subjects needed to drag the slide from the current position/floor to the requested floor level, or press the up/down button to make the slide move two floors. The stab control was operated in a manner similar to the scroll bar control.

Task 3. Check floors to find the fire. Subjects were told that the fire went off between Floor A and Floor B, and were asked to find it. For all three vertical navigation methods, one way to find the fire was to press the up/down button to check each floor until the fire was found. For the scroll bar and stab control an additional method was to drag the slide or current floor card smoothly, until the fire was found. For the soft keyboard display an additional method was to delete the current floor number and input a new one to find the fire.

3 Procedure

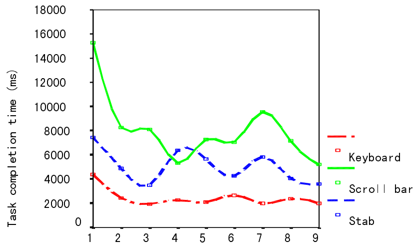
Without any training, subjects were asked to accomplish the three experimental tasks one by one. The order of the three tasks was counter-balanced by Latin square between subjects. For each task, subjects were asked to accomplish by the following three ways respectively and randomly: keyboard, scroll bar, and stab control. Each part included 20 trials. After a total of 60 trials, a short subjective evaluation about the three kinds of vertical displays was conducted. During the test, screen resolution was set at 1280 by 1024.

Participants: Since all the firefighters in China are male soldiers, and their age is around 18 to 36 years, 14 male subjects at the same age participated in this experiment. Each subject took half an hour to accomplish the three experimental tasks. This was followed by a 5-minute interview.

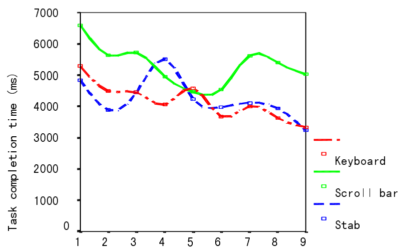
4 Results

4.1 Practice Effect

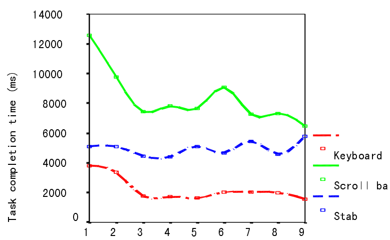
The interface of the Fire Information Display System should not require any training. So in this study, subjects didn't get any training, or introduction before the test. They had to guess how to accomplish the tasks. Subjects had 20 trials of each task. The task completion time and the number of control operations were recorded as their performance. Since there were some data lost at the 10th and the 20th trial, only the first 9 performances were analysed.



(a) Task 1. Move to a specific floor



(b) Task 2. Switch to another floor



(c) Task 3. Check floors to find the fire

Fig. 3. Practice effects on the three experimental tasks

Fig. 3 (a)- (c) shows the different practice patterns when people use soft keyboard, scroll bar, or stab control as the vertical navigation method to accomplish the three experimental tasks. For Task 1 (finding a specific floor as directed; see Fig. 3 (a)), the first trial took significantly longer than the other 8 trials ($F(2, 8) = 4.196$, $p < .001$). The task completion time (TCT) using the soft keyboard kept stable around 2 seconds, and the TCT for the scroll bar and stab control fluctuated between 8 seconds and 4 seconds. However, these differences across the eight trials were not significant. In other words, people needed only one trial to learn how perform Task 1 with all three kinds of floor selection methods.

For Task 2 (switching from current floor to two floors higher/lower; see Fig. 3 (b)), there was no difference across the 9 trials. This means that after the Task 1 was completed, Task 2 didn't need any further training to use soft keyboard, scroll bar and stab. But the TCT of all three vertical navigation methods fluctuated between 6 and 4 seconds.

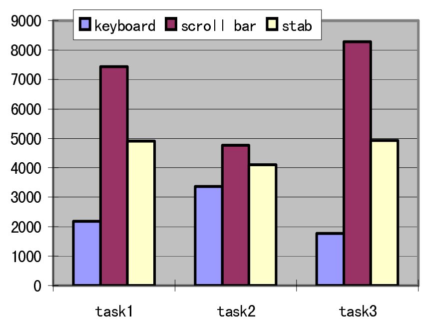
For Task 3 (check floors and find the fire; see Fig. 3 (c)), the TCT of using soft keyboard still was stable at 2 seconds, TCT of stab at 5 seconds. But on the first two trials, the TCT of using the scroll bar was significantly longer than the TCT of the other 7 trials. The amount of training required for the scroll bar in Task 3 was two times that for the other methods.

4.2 Task Completion Time

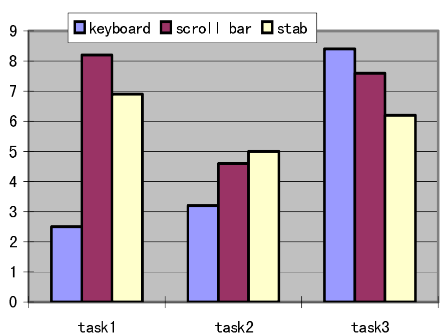
Since after the first two trials, there was no significant difference among the other trials, the mean of TCTs in the last 16 trials was calculated as each subject's performance. The

Table 1. Mean task completion times (msec.) for three vertical navigation controls

Task	Vertical navigation			p
	Soft keyboard	Scroll bar	Stab	
1	2181.9	7440.8	4907.9	F(2)=93,p<.01
2	3362.4	4760.6	4096.4	F(2)=6.4,p<.01
3	1772.7	8284.5	4925.1	F(2)=83.7,p<.01
p	F(2)=11.8,p<.00	F(2)=23,p<.01		



(a) Task completion time



(b) numbers of screen operations during each trial

Fig. 4. Comparison of three vertical navigation controls on different tasks

average TCT under different experimental situations are listed in Table 1. All the TCT ranged from 1700ms to 8300ms. Based on the ANOVA for repeated measures, the scroll bar display made the TCT the longest in all three tasks, and the soft keyboard made it the shortest (F values and p values are shown in Table 1). The TCT under stab situation kept stable at 4000ms to 5000ms (also see Fig.4 (a)). This result was just opposite to the hypothesis we made, which probably because that people are much more familiar with keyboard than with scroll bar and stab. Scroll bar and stab still were quite new things for subjects. Especially people didn't use them to input information to computer system. In addition, when using scroll bar to locate a position, people usually took two steps: first to drag the slide to a position roughly, then move the slide very slow and adjust the position very carefully, or use up/down button to adjust the position with very small steps. But in this study, it's found that some subjects didn't use up/down button, it took subjects long time to drag the slide and adjust the position.

4.3 Number of Screen Operation

Numbers of screen operations were recorded as another criterion on which to compare the three vertical navigation methods. Sometimes people took a shorter time to input a floor number, or switch floors, but the way of vertical navigation made people touching the screen button or moving their finger on the screen many times to accomplish a simple task. It didn't mean easy and intuitive.

Table 2 lists the average number of screen operations under different experimental conditions. The soft keyboard cost the least in operation time on Task 1 and Task 2. There was no difference in number of operations between the stab control and scroll bar. The keyboard was always the best way to input information accurately, and the scroll bar and stab control needed more individual input operations to adjust their positions very carefully. But for the Task 3, which involved checking floors to find the fire, the soft keyboard lost its advantage. The number of operations for the keyboard method increased dramatically, owing to the requirement to change floors very frequently. In addition, some subjects didn't notice the left and right arrow buttons that could have helped them change floors very quickly. The scroll bar and stab control were not sensitive to different tasks including Task 3.

Table 2. Numbers of screen operation

Task	Vertical navigation			p
	Soft keyboard	Scroll bar	Stab	
1	2.5	8.2	6.9	F(2)=4.65, p=.014
2	3.2	4.6	5.0	F(2)=4.09, p=.025
3	8.4	7.6	6.2	No difference
p	F(2)=38., p<.00	No difference	No difference	

4.4 Subjective Evaluation

After the test, 16 subjects accepted a 5-minute follow-up. They were asked to answer the following questions:

1) For Task 1, which way do you think is the best way to input the floor number and get to the floor? Which one is the worst? Why?
2) For Task 2, and Task 3, which way do you think is the best way to check floors and find the fire? Which one is the worst? Why?
3) Did you use the up/down arrows during the test?
4) Are there some features you think could be improved?

For questions 1 and 2, the most preferred navigation method received a score of 3 and the least preferred received a score of 1. The average scores for the three methods of vertical navigation are shown in Table 3. It showed that people did not prefer any one of the vertical navigation methods to another on tasks 2 and 3. But they did prefer using the soft keyboard to input a floor number in Task 1.

In addition, five subjects said they didn't notice there were up/down arrow buttons at the two ends of scroll bar and stab control. Another six subjects didn't notice that there were left/right arrow buttons on the soft keyboard interface. Up/down buttons were critical for people to adjust the current position by one floor. The result showed that the up/down buttons were not apparent enough to draw people's attention. Some subjects suggested making the size of up/down buttons a little bigger than they were, or make their colour different from other buttons.

Most people liked the scroll bar display, but felt that the current implementation was not good enough. For example, the slide was not sensitive enough to the finger's dragging movements. Subjects always complained that the slide didn't follow their finger going down/up quickly and smoothly. It made them feel frustrated. It probably was the main reason why the scroll bar display didn't show more advantages than the soft keyboard. This suggests that we improve the current design and conduct another experiment in the future.

People liked the stab control too, but they felt the interval between two floors was too narrow for their fingers to distinguish.

Table 3. Subjective evaluations of three kinds of vertical navigation controls

Task	Vertical navigation			p
	Soft keyboard	Scroll bar	Stab	
1	2.5	1.7	1.8	F(2)=4.8, p=.026
2,3	1.9	1.9	2.1	No difference

5 Discussion and Conclusion

In this experiment, the soft keyboard was superior to the scroll bar and stab controls for vertical navigation of a high rise building on a touch screen. We suspect that there were two reasons for this:

1. With no training on the three navigation controls prior to performing the high rise navigation tasks, the keyboard simply was the most familiar way for

subjects to input information. Everyone, including people who lack of computer experience, and know how to use a keyboard. It has been standardized and is widely accepted.

2. The scroll bar design used in the study needs to be improved. The implementation used in the experiment was not sensitive enough for a touch screen. In particular, the slide in the scroll bar proved to be not big enough. A slide bigger than 50 X 20 pixels needs to be designed and tested. The Stab control suffered from similar problems.

Although the scroll bar and stab controls produced lesser performance than the keyboard for high rise vertical navigation, the subject's comments indicated that, with some re-design, these might be valid alternatives to the keyboard for touch screens. It should be noted that the scroll bar and stab control could be used as displays also. That is, they could provide an intuitive way to indicate the currently selected floor position in a high-rise building. Further exploration of this is left for the future study.

Acknowledgments. Thanks to the people from Beijing Fire-Fighting Bureau who gave us access and helped us in this study.

References

1. Aftab, E.P, Sebastian, S., Tomescu, G.A., M.: Ishac: What visual information is used for navigation around obstacles in a cluttered environment? Canada Journal of Physiology and Pharmacology 82, 682–692 (2004)
2. Vries, d., Bauke, Buma, Sjoerd, Jessurun, Joran.: An Intuitive Interface for Building Management and Planning. International Journal of Architectural Computing 4(2), 17–26 (2006)
3. Feary, M., Alkin, M., Polson, P., McCrobie, D., Sherry, L., Palmer, E.: Aiding vertical guidance understanding. Air. & Space Europe 1(1), 38–41 (1999)
4. Fraser, Peter, J.: Review: Depth, navigation and orientation in crabs: Angular acceleration, gravity and hydrostatic pressure sensing during path integration. Marine and Freshwater Behaviour and Physiology 39(2), 87–97 (2006)
5. Yao, G.: Application of touch screen in HCI. Construction Machinery & Maintenance 10, 16–20 (2004)
6. Zhang, Y., Zhang, Z., Zhang, J., Wu, J.: 3D Building Modelling with Digital Map, Lidar Data and Video Image Sequences. The. Photogrammetric Record 20(111), 285–302 (2005)

Mental Workload in Command and Control Teams: Musings on the Outputs of EAST and WESTT

Mark S. Young, Neville A. Stanton, Guy H. Walker, Daniel P. Jenkins, and Paul M. Salmon

School of Engineering and Design, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK
m.young@brunel.ac.uk

Abstract. The EAST methodology (Event Analysis of Systemic Teamwork) has recently been developed as a means for analyzing command and control scenarios. Its counterpart, the WESTT computer-based analytical prototyping tool (Workload, Error, Situation awareness, Time and Teamwork) allows analysts to visualize, model and measure these command and control networks and activities. Recent studies applying EAST and WESTT to the military domain have led to simplistic inferences about the relative levels of demands during phases of task activity. The present paper takes these inferences a step further, suggesting alternative methods for evaluating mental workload based on human information processing models and network metrics such as latent semantic analysis. Whilst much development work is implied, these analyses offer the possibility not only to quantify mental workload across the network, adding further value to the EAST/WESTT toolkit.

Keywords: mental workload, command and control, teamwork, network analysis.

1 Networks in Command and Control

Research within the Human Factors Integration Defence Technology Centre (HFI DTC) has led to the development of the EAST methodology (Event Analysis of Systemic Teamwork; Walker et al., 2006) for analyzing command and control scenarios. EAST is a concatenation of several ergonomics methods in order to represent three key aspects of complex socio-technical systems: the task network (how the task is structured), the social network (who is doing what), and the information network (what are the data requirements of the task). In parallel, an analytical prototyping software tool called WESTT (Workload, Error, Situation awareness, Time and Teamwork; Houghton et al., in press) has been developed to allow analysts to visualise, model and measure these command and control networks and activities. Thus EAST and WESTT are complementary techniques in analysing a complex team-based structure from both retrospective and predictive viewpoints.

Both EAST and WESTT have been successfully applied to a range of command and control scenarios, with the main focus of their development being in the military domain (e.g., Walker et al., in press). Recent studies (Walker et al., in press) have

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 455–464, 2007.

applied both techniques to an assessment of the military's main planning process, the Combat Estimate. Using the output of the information networks, it is possible to identify key information objects which are 'active' during each phase of the planning process. Walker et al. (in press) then summarise these core information objects to imply the relative levels of demand during each phase.

In broad terms, then, the amount of information used in the system at any one time has been taken as a rudimentary measure of mental workload (MWL). Such an approach may be useful for gaining an overview of the task, but a more thorough evaluation will be necessary if the methods are to be used in detailed workload evaluations. MWL is a multidimensional construct, comprising elements of task demands (in turn set by performance criteria), operator skills and experience, and any external support from technological or human team members (cf. Young & Stanton, 2001). A simple tally of the active information in the network does not necessarily reflect such qualitative aspects of that information (i.e., task difficulty), nor does it account for the members of the team who are involved on that task.

The EAST and WESTT methods do, however, offer potential for exploring these aspects of MWL in detail. The task networks provide information about what needs to be carried out – that is, the objective performance criteria against which task success or failure will be judged. The social networks can inform us about who is performing the task, thus mediating the workload on individuals by means of team support and/or an assumed level of skill. Finally, the information networks offer much promise in terms of qualitative and quantitative aspects of task demands. The overall network has a number of statistical properties regarding information flow and communication, and which also suggest the 'importance' of individual elements.

The present paper explores the potential for developing EAST and WESTT to assess MWL in the command and control system. We begin with an overview of each method and their applications, before going on to discuss the possibilities for linking networks in order to cross-reference tasks, teams and information requirements. Finally, we examine alternative discourse analysis methods such as latent semantic analysis (Landauer et al., 1998) as methods of capturing information on the salience or complexity of information elements.

1.1 EAST

The EAST methodology was developed to feed into generic models and theories of C4i (command, control, communications, computers and intelligence; see Harris & White, 1987), in order to optimize the effectiveness of C4i scenarios (Walker et al., 2006). EAST is a retrospective analysis, integrating multiple methods from the human factors literature base (Stanton et al., 2005) into one methodology to describe, analyze and represent multiple aspects of C4i activity. This approach allows the data to be analyzed from multiple perspectives, and helps to establish the *why, who, when, where, how* and *what* of these complex systems.

In terms of the outputs from EAST, these essentially comprise three types of network analysis (Stanton et al., 2005): the *task* network (*what* is being performed, *why* is it important, and *how* is the task structured), the *social* network (*who* is performing the tasks), and the *information* network (*what* information is required for the tasks to be completed effectively). Task networks are derived from Hierarchical

Task Analysis (HTA, Annett, 2003); social networks from Social Network Analysis (SNA, Driskell & Mullen, 2005); and information networks from propositional network analysis (cf. Ogden, 1987). Arguably the most novel and potentially useful output from the EAST methodology, propositional networks are analogous to semantic networks in that they contain nodes (with words) and links between nodes. Nodes in this context represent the necessary and required 'information elements' for the system to operate effectively, and are independent of objects, people or ideas (Walker et al., 2006). Since these nodes specify the information requirements of the system, and which 'agents' possess such knowledge, the propositional network has been used as a new way of evaluating distributed situation awareness (DSA) in systems (Stanton et al., 2006). Figure 1 displays an example propositional network from part of a railway maintenance task (Walker et al., 2006):

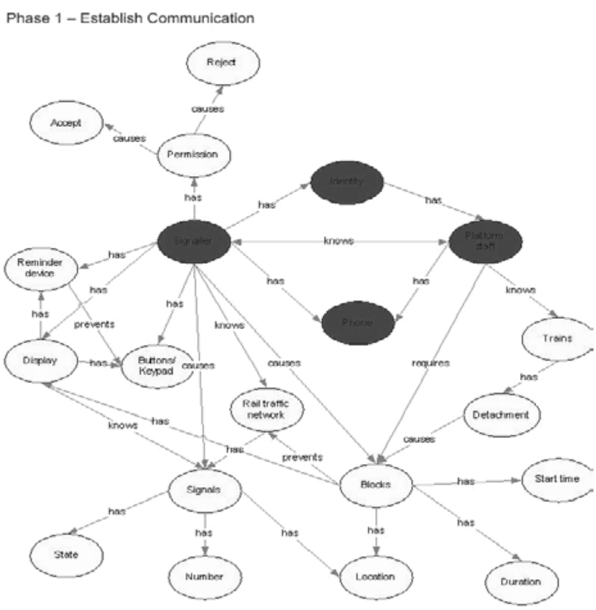


Fig. 1. Example propositional network detailing a railway maintenance task (from Walker et al., 2006)

Note how each node is simply a piece of information, which is related to other nodes by relevant propositions (e.g., 'has', 'causes', 'requires'). As such, information flow around the system can be modeled, and nodes can become 'active' as the task sequence progresses (in figure 1, active nodes for the task phase have been shaded).

In EAST, the propositional networks are populated by data collection from the Critical Decision Method (CDM; Klein & Armstrong, 2005), a specific and detailed form of interview. The transcripts of subject matter experts are used to extract the information elements and the propositions linking them. Since the relevant knowledge is based on key decision points, the propositional networks provide an opportunity to describe all the key elements related to that decision point.

As a representation of the 'ideal' knowledge for any given scenario, propositional networks can be useful in determining the overall information requirements of the system. However, the data in a propositional network can be subjected to more quantitative analysis, and that is where the WESTT software tool comes in.

1.2 WESTT

Where EAST is a retrospective analysis, the WESTT (Workload, Error, Situation awareness, Time and Teamwork) is a predictive human performance software modeling tool (Houghton et al., in press), developed to represent, analyze and model team activity in complex C4i scenarios. WESTT represents system activity according to the information elements (i.e., propositional networks) and the agents who carry out the activities (i.e., social networks). Its primary inputs are data from field observations, which can also be automated using a hand-held PDA device, supplemented by interviews with subject matter experts. As such, WESTT is an ideal software tool to support the EAST methodology.

WESTT populates an internal database, organized chronologically by task, agents, and communications, from which the tool can automatically build the task, social and information networks familiar from EAST. In addition to these qualitative outputs, WESTT allows a certain level of quantification of these networks.

At one level, WESTT can total up the active tasks, time-on-task, and activities by agent, allowing an assessment of team and operator workload. But in perhaps a more promising area of research, the social and propositional networks can be analyzed using statistical techniques adapted from graph theory. This is a growing area of mathematical interest and provides various measures of positional centrality (such as sociometric status) which relate to the level of interconnection each individual node has, which, by implication, might relate to that node's importance. In addition, it is possible to calculate the overall density of the network, which indicates the level of interconnectivity between agents (Walker et al., 2006). That is, the ratio of network links being used, compared to the maximum theoretically available.

In terms of social networks, these measures can be interpreted as indicators of the key agents in the scenario. Both sociometric status and centrality are effectively measures of power or influence in a network, based on the substantive and structural contribution of the node to the network relative to other nodes (Friedkin, 1991). Similarly, the information elements of propositional networks can be subjected to the same analyses, and thus trends in the relationships between information can be identified. Since the centrality measures indicate key information objects, it becomes possible to identify the underpinnings of situation awareness in the system at a particular phase in the task. WESTT even takes this a step further, linking the social and information networks to represent possession of knowledge – 'who knows what' – during a phase in the task.

Whilst WESTT offers a promising way forward for assessing potential errors, situation awareness (via system structure, distribution of information objects etc.), and teamwork (e.g., the role of human vs. non-human agents in the system), a question remains about task demands within the network – that is, the mental workload imposed by the system.

2 Mental Workload

Mental workload (MWL) has been traditionally and popularly conceptualized in terms of a task-timeline (e.g., Parks, 1979). In other words, workload estimates are derived by analyzing the range of tasks to be performed and plotting these against time. The timeline can then help to identify peaks in MWL where multiple concurrent tasks are necessary, or where the time required to perform tasks approaches or exceeds the maximum time available. This simplistic ratio can provide an indication of demands, conflicts and overload, and essentially forms the basis of the workload estimator within WESTT.

However, many authors have noted that MWL is a complex, multidimensional concept, and agreement on what constitutes MWL, or indeed how it can best be quantified, remains elusive in the human factors and ergonomics literature. The level of MWL imposed upon an individual results from facets of the task (e.g., demands, performance) and of the operator (skill, attention). This prompted Young & Stanton (2001) to propose a broad-ranging definition of MWL, as the "…level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support, and past experience" (p. 507). The one-dimensional timeline models seem to be crude in the light of this multifaceted approach to the underlying construct. Since timeline models view the human operator as a single-channel, limited capacity processor, they do not take account of factors such as multiple attentional resources (Wickens, 2002), qualitative task intensity (cf. Hockey, 1997), or operator skills (Gopher & Kimchi, 1989).

Hendy et al. (1997) demonstrated support for a model of human information processing which reconciled these time- and intensity-based models of MWL. They showed that in fact, a time ratio model built on task networks determined both subjective ratings of MWL as well as operator performance. Whilst this would seem to support the simplistic approaches such as incorporated in WESTT, the model put forward by Hendy et al. (1997) was rather more complex, and was based not merely on the time required to complete the task, but the rate of information processing demand (RID). Their time pressure ratio then becomes RID divided by channel capacity. Thus it is still necessary to know something of information requirements as well as the capacity of the operator.

The picture becomes even more complex when considering MWL at the team, rather than the individual, level. Indeed, this is an area which has been comparatively neglected in the literature (Bowers et al., 1997). In their review, Bowers et al. (1997) note that the coordination and communications activities associated with teamworking itself impose demands over and above the actual taskwork to be achieved. It is conceivable, of course, that such teamworking activities reduce the taskwork by means of 'external support' (cf. Young & Stanton, 2001). Under a sociotechnical systems perspective, such support can be in the form of other human team members, or from technological means (e.g., automation). This perspective is consistent with the network approach of referring to 'agents' in the system, allowing tasks, information (and hence MWL) to be distributed across the team. It should be possible, therefore, to predict and measure team MWL on the output of these propositional networks. Measures of team MWL have thus far, however, been restricted to subjective evaluations (see Stanton et al., 2005). Since we have seen that the propositional

network approach offers a promising avenue for exploring DSA, it seems intuitive that we could equally use similar network metrics to evaluate distributed MWL.

Thus we have identified a clear requirement for a more robust and quantitative approach to MWL measurement in the task, social and information networks such as output from the EAST and WESTT methods. The remainder of this paper explores possible avenues for developing such metrics.

3 Distributed MWL

Thus far, we have noted that the propositional networks as output by EAST and WESTT offer a rich source of data about information and communication requirements in the system to support effective performance. However, whilst progress is being made on using the propositional networks to advance theories of DSA, the MWL analyses offered so far have been crude in comparison.

In WESTT, for instance, the software creates Operational Sequence Diagrams (OSDs, another output of the EAST methodology) to graphically represent activities and agents over time in a flowchart style. These activities can then be tallied up as an assessment of team workload, and estimates of task duration and time-on-task can be generated (Houghton et al., in press). These OSDs are very reminiscent of the timeline analyses of MWL previously referred to; it is indeed a short step to calculate the ratio of time required against time available as advocated in early MWL literature (e.g., Parks, 1979).

Similarly, Stanton et al. (2006) speculate that workload across the team will be a function of the activated core information objects at any particular phase in the task. They further define core information objects according to an arbitrary cut-off point on the network measures of sociometric status and centrality. A simple count of core information objects is suggested as being correlated with team workload (see Figure 2). Whilst this may be true to a certain extent, such an analysis takes no account of the qualitative nature of the demands (i.e., the task intensity), nor of any time pressure requirements. Furthermore, additional load arising from teamworking activities is not addressed, although Stanton et al. (2006) intimate that information communication is an important factor in the network structure.

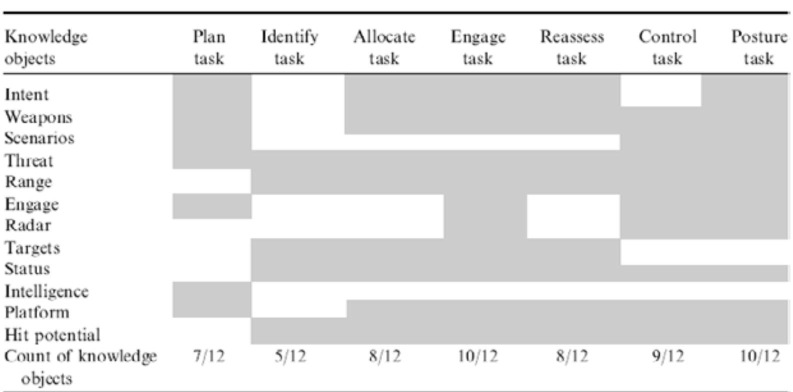


Fig. 2. Analysis of core information objects during Royal Navy frigate operations (from Stanton et al., 2006)

Rather, we feel that a more sophisticated approach is necessary. Taking elements from Hendy et al.'s (1997) information processing model, coupled with further use of network approaches, it is posited here that EAST and WESTT can offer a more detailed analysis of MWL across the network – that is, of *distributed MWL*.

As discussed previously, the model proposed by Hendy et al. (1997) reconciles time-based and intensity-based models by considering the rate of demand – that is, the information flow rate – against information processing capacity. Whilst they offer convincing evidence that this measure of time pressure is an accurate reflection of MWL and predictor of performance, it remains difficult to see how it may be extrapolated to other situations without *a priori* definitions of flow rate and capacity. The propositional networks of EAST and WESTT, however, do contain such data within the nodes and the propositions between them; the only problem is accessing it in the right way. The next step is therefore to relate these properties to the salience and complexity of the information contained within the elements – in other words, a way of 'tagging' each element with a difficulty factor.

The network measures used so far in the development of EAST and WESTT – sociometric status and centrality – are not entirely appropriate for this difficulty factor. Whilst they do provide some indication of the importance of the node in the network, this relates more to the awareness contained within the system and the transmission of information around the network. Delete a key information element, and the rest of the network will either have to adapt (which will, of course, have workload implications), or performance will suffer. However, this is not the same as rating the complexity of the information within the node. We still need that measure of task intensity, or rate of information processing demand.

Given that we are addressing network structures, it seems sensible to appeal to the original cognitive psychological research on semantic networks, connectionist structures, and parallel distributed processing (e.g., Anderson, 1983; McClelland, et al., 1986). These models refer to levels of activation within the nodes in order for information to 'fire', as well as the strengths of links between nodes which can facilitate or inhibit information transfer.

Recall the systems view of DSA (Stanton et al., 2006), which can be described as activated information at the network level (Bell & Lyon, 2000), since the nodes become active at different phases of the task. It is worth noting that, for individuals at least, MWL and SA have been related, inasmuch as high MWL can disrupt SA as it absorbs attentional resources (Stanton & Young, 2000). In other words, MWL could inhibit activation of the nodes. Conversely, the activation level of a particular node could feasibly represent its level of information processing demand, its MWL.

In our propositional networks, then, the activation level could be a quantifiable analogue to the complexity of the information – that is, the task intensity in terms of information bits per second – while the connection strength is essentially the bandwidth – that is, the channel capacity. If we know the time available for performing the task (from the task network or the OSD), then after Hendy et al. (1997), we can derive the rate of information processing demand as bit rate divided by time available. Further, this subsequently allows us to predict the time pressure as rate of information processing demand divided by channel capacity.

Although Hendy et al.'s (1997) model is clearly designed for an individual information processor, given the propositional networks here for a sociotechnical

system, there is no logical reason why the derived model cannot be applied in a distributed sense. The next problem, then, is defining and quantifying activation level and connection strength.

For connection strength, EAST already offers an answer with the Communications Usage Diagram (CUD; Watts & Monk, 2000). Channel capacity for the individual is a function of processing resources (Hendy et al., 1997); for the team (or system) it would be a function of communication resources. Whilst the CUD only provides a qualitative assessment of communications media, the number of channels available and a measure of their efficiency could be used to evaluate connection strength.

Bit rate, or task intensity, presents a rather more difficult problem. How to model the complexity contained within a node is a challenge which goes beyond mere salience (i.e., measured with centrality). An information element can be core to the system, yet remain a relatively simple piece of information. Instead, we need a way of modeling the absolute processing demands involved with that information. Again, we appeal to the wider literature on semantic networks for a solution.

Latent semantic analysis (LSA; Landauer et al., 1998) can be used as a "…model of the computational processes and representations underlying substantial portions of the acquisition and utilization of knowledge". LSA is essentially a form of discourse analysis, which uses statistical procedures similar to factor analysis in order to extract and represent the contextual-usage meaning of words. As such, then, it can be applied to the transcripts collected from the CDM in the EAST process. Using LSA in this way allows some measure of inductive learning, and hence an assessment of the information quality available in the system. It follows that if information is impoverished, MWL will be increased. The inductive reasoning abilities of LSA have been shown to distinguish between easy and difficult questions on university exams, as well as distinguishing between expert and novice knowledge on a given subject (see Landauer et al., 1998). This evidence supports the potential of LSA to evaluate the processing demands associated with the information network. As Landauer et al. (1998) state, "…the success of LSA as a theory of human knowledge acquisition and representation should … not be underestimated".

If the outputs of LSA can be adapted to provide some quantification of information complexity in the network, and if CUD does provide appropriate measures of channel capacity, then with some further development, EAST and WESTT can be used to derive a metric of distributed MWL. We use the term 'distributed MWL' deliberately as a corollary of DSA, since it represents a qualitatively different, more integrated concept than overall team MWL. Team MWL purports to distinguish between taskwork and teamwork; distributed MWL is more about the acquisition, processing, and management of information around the network. In terms of the information processing model (Hendy et al., 1997), distributed MWL is a direct correlate of individual MWL. From this perspective, MWL can be altered either by changing the rate of information processing demand (i.e., the nature of the task in terms of its information content, or time available), or the channel capacity (through communications media or additional team agents). Practical recommendations for team and work design can then be made, all of which are potentially available already in the task, social and information networks of EAST and WESTT.

4 Conclusions

Clearly there is much work still to be done in developing the appropriate metrics for both level of activation (via LSA) and connection strength (via CUD) in order to calculate the rate of information processing demand (cf. Hendy et al., 1997). Within the scope of this paper, we could not – and did not intend to – develop such metrics. Rather, as the title suggests, these are musings on the potential for metrics of distributed MWL in complex networks, such as the C4i scenarios to which EAST and WESTT have been applied. Future work could seek to develop links between the task, social and information networks, in order to quantify distributed MWL from a multidimensional perspective. Ultimately, a detailed and structured analysis of distributed MWL would be an invaluable addition to the EAST/WESTT toolkit, as the understanding of MWL distribution amongst the team and across the task would allow us to optimize the team design process according to tasks, resources and people (cf. MacMillan et al., 1999).

References

1. Anderson, J.R.: The Architecture of Cognition. Harvard University Press, Harvard (1983)
2. Annett, J.: Handbook of Task Analysis in Human-Computer Interaction. In: Diaper, D., Stanton, N.A. (eds.) Handbook of Task Analysis in Human-Computer Interaction, Lawrence Erlbaum Associates, Mahwah, NJ (2003)
3. Bell, H.H., Lyon, D.R.: Using observer ratings to assess situation awareness. In: Endsley, M.R. (ed.) Situation Awareness Analysis and Measurement, pp. 129–146. Lawrence Erlbaum Associates, Mahwah, NJ (2000)
4. Bowers, C.A., Braun, C.C., Morgan, B.B.: Team workload: Its meaning and measurement. In: Brannick, M.T., Salas, E., Prince, C. (eds.) Team Performance Assessment and Measurement: Theory, Methods and Applications, pp. 85–108. Lawrence Erlbaum Associates, Hillsdale, NJ (1997)
5. Driskell, J.E., Mullen, B.: Social Network Analysis. In: Stanton, N.A., et al. (eds.) Handbook of Human Factors and Ergonomics Methods, vol. 58, pp. 1–58.6. CRC, London (2005)
6. Friedkin, N.E.: Theoretical foundations for centrality measures. The American Journal of Sociology 96(6), 1478–1504 (1991)
7. Gopher, D., Kimchi, R.: Engineering psychology. Annual Review of Psychology 40, 431–455 (1989)
8. Harris, C.J., White, I.: Advances in command, control & communication systems. In: IEE Computing Series 11, Peregrinus, London (1987)
9. Hendy, K.C., Liao, J., Milgram, P.: Combining time and intensity effects in assessing operator information-processing load. Human Factors 39(1), 30–47 (1997)
10. Hockey, G.R.J., Maule, A.J.: Unscheduled manual interventions in automated process control. Ergonomics 38(12), 2504–2524 (1995)
11. Houghton, R. J., Baber, C., Cowton, M., Stanton, N. A.: WESTT (Workload, Error, Situational Awareness, Time and Teamwork): An analytical prototyping system for command and control. Cognition, Technology and Work (in press)

12. Klein, G., Armstrong, A.A.: Critical decision method. In: Stanton, N.A., Hedge, A., Brookhuis, K., Salas, E., Hendrick, H. (eds.) Handbook of Human Factors and Ergonomics Methods, vol. ch. 58, CRC Press, Boca Raton, Fl (2005)
13. Landauer, T.K., Foltz, P.W., Laham, D.: An introduction to latent semantic analysis. Discourse Processes 25, 259–284 (1998)
14. MacMillan, J., Paley, M.J., Serfaty, D., Levchuk, Y.N.: Designing the best team for the task: A method that combines algorithms, heuristics and expert judgment. In: Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting, Houston, TX (1999)
15. McClelland, J.L., Rumelhart, D.E., Hinton, G.E.: The appeal of parallel distributed processing. In: Rumelhart, D.E., McClelland, J.L. the PDP Research Group (eds.) Parallel Distributed Processing, MIT Press, Cambridge, MA (1986)
16. Ogden, G.C.: Concept, knowledge and thought. Annual Review of Psychology 38, 203–227 (1987)
17. Parks, D.L.: Current workload methods and emerging challenges. In: Moray, N. (ed.) Mental Workload: Its Theory and Measurement, pp. 387–416. Plenum Press, New York (1979)
18. Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C., Jenkins, D.P.: Human Factors Methods: A Practical guide for Engineering and Design. Ashgate, Aldershot (2005)
19. Stanton, N.A., Stewart, R., Harris, D., Houghton, Baber, C., McMaster, R., Salmon, P., Hoyle, G., Walker, G., Young, M.S., Linsell, M., Dymott, R., Green, D.: Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. Ergonomics 49(12-13), 1288–1311 (2006)
20. Stanton, N.A., Young, M.S.: A proposed psychological model of driving automation. Theoretical Issues in Ergonomics Science 1(4), 315–331 (2000)
21. Walker, G., Gibson, H., Stanton, N., Baber, C., Salmon, P., Green, D.: Event analysis of systemic teamwork (EAST): A novel integration of ergonomics methods to analyse C4i activity. Ergonomics 49(12-13), 1345–1369 (2006)
22. Walker, G.H., Stanton, N.A., Stewart, R., Jenkins, D., Wells, L., Young, M., Green, D., McMaster, R., Baber, C.: Using an integrated methods approach to analyse the emergent properties of military command and control. Cognition, Technology and Work (in press)
23. Watts, L.A., Monk, A.F.: Reasoning about tasks, activities and technology to support collaboration. In: Annett, J., Stanton, N. (eds.) Task Analysis, pp. 55–78. Taylor & Francis, London (2000)
24. Wickens, C.D.: Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science 3(2), 159–177 (2002)
25. Young, M.S., Stanton, N.A.: Mental workload: theory, measurement, and application. In: Karwowski, W. (ed.) International encyclopedia of ergonomics and human factors, vol. 1, pp. 507–509. Taylor & Francis, London (2001)

Lightweight Collaborative Activity Patterns in Project Management

Shaoke Zhang[1], Chen Zhao[1], Paul Moody[2], Qinying Liao[1], and Qiang Zhang[1]

[1] IBM China Research Lab, Bldg 19, Zhongguancun Software Park,
Beijing 100094, P.R. China
shaoke_zhang@hotmail.com, {zhaochen, liaoqy, zhqiang}@cn.ibm.com
[2] IBM Cambridge Lab, One Rogers Street, Cambridge,
MA02142, USA
paul_moody@us.ibm.com

Abstract. People working with ad-hoc collaboration tools suffer from information overload and information scatter. Our five-month study of project managers found their work comprised of fragmented activities scattering in task, people and application. However, these loosely coupled collaborative activities were implicitly organized by hierarchical activity threads through intrusion and digression, and geared into work breakdown structure (WBS) in high level. Our investigation gives implications on making people's work manageable as well as flexible, by integrating lightweight collaborative activities with centralized processes.

Keywords: Fragmented activity, activity threads, project management.

1 Introduction

Now increasing researchers have noticed people are suffering from frequent interruptions, overwhelming information, and high workload in work [7] [9] [16]. The development of information and networking technologies greatly accelerates this trend by facilitating various collaborative activities simultaneously. However, people's extremely limited cognitive capabilities cannot catch up with the technological advances, which leaded to inefficiency, stress, annoyance, and anxiety [1]. Terms like "the plague of managerial work" and "time famine" usually appear in studies stating their stubborn status [9]. And issues like the confined short-term working memory, long-term working memory, and prospective memory are also increasingly discussed in these problems [6] [15].

People's work activities have been supported in flexible and lightweight ways by various ad hoc tools like instant message, email, todo lists, office tools, etc. All these tools are designed to help the user perform particular tasks to completion, and thus have always been maintained separated. While our work are mentally organized by activity threads (or thrasks, working spheres) [2] [7] [11], they practically cut across these tools, which can introduce frequent switches and interruptions, and thus high workload, unmanageable information, and low productivity [16]. Moreover, to response to the increasing functionally integrating requirements, rather than incorporating these tools

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 465–473, 2007.

together, these tools tend to include more and more features and lead to many function overlaps and make people's work even more confused.

Therefore, a number of researchers are attempting to support and organize collaborative work around activities [2] [5] [8] [11] [12] [13] [14]. While there have been efforts to manage activities in specific tools like email and to-do lists [3] [4], many other researchers are trying a new approach to incorporate various tools together. As early as 1980s, Bannon et al. and Card et al. [2] [5] claimed that there should be a "workspace" that integrated various task systems. And now there is a trend of studies concerning with topics like Unified Activity Management [12], Activity-Centric Approaches [8] [11] [13], and shared activities [14]. The Activity Explorer was such a tool developed to support shared objects like chat, email, document etc.; and these objects were organized by activity threads [11] [13]. These studies have the notion that technology should support people focusing on the work they are trying to accomplish [12].

However, the foundation of this kind of support is an explicit representation of people's work, which has not been sufficiently and systematically understood. In this paper, we investigated project manager's work activities to have a better understanding of the lightweight collaborative activity patterns, which would provide implications for how on earth these loosely coupled activities should be supported. We selected project management (PM) as the study scenario because PM work typically comprised intensive and complex lightweight collaborative activity. Besides, rather than just giving some statistics, the projects could serve as cases with rich context of the activities, which helped us to meaningfully understand the causes, lifecycles, interdependencies, and effects of the activities, and to grasp how the activities patterns were really like.

2 Method

A combination of ethnographic techniques including interview and field observation was used in this study. We first interviewed 17 project managers or project leaders separately. These semi-structured, in-situ interviews lasted about 60-90 minutes, providing information on project management activities, processes, tools usage, and problems. The data consisted of participant profiles, project descriptions, interview notes, audio recordings, photographs of office settings, and samples of artifacts. Interviews took place in their offices during their work time, which provided context cues for the interview. Our participants' industries included software development, banking, construction, and IT service. Their project management experience ranged from 1.5 to 16 years, with an average of 5.92 years (*SD* = 4.01). Their average ages was 33.58 years (*SD* = 4.91). The size of project teams varied from 5-6 members to more than 40 members.

Based on the understanding acquired from interview, we further conducted field study of another two projects in a more tangible and quantitative way. Both these two projects were run at customer sites (one in Beijing, the other in Tianjin) where our

observations took place. The three-month long observation occurred in the initiation & engagement phases in one project, and in the closing phase in another. Spotlighting on two selected project managers, we observed their work activities for about eighty one hours, including their collaboration with other forty five persons.

The work settings in both two projects were open; people including the customer deputies, project members, and vendors sat in the neighboring offices. While we originally followed project managers all day as he or she worked in the office, (like the “shadowing” observation technology used in [7]) we also captured other people’s communication with project managers. At least two researchers participated in each visit. One researcher took notes, while the other video recorded interactions with people and computer in the office and meeting room. The researchers occasionally asked questions during breaks without interrupting. At the end of each day, we had a brief retrospective interview, and asked participants to clarify some questions. We collected physical and electronic materials and took pictures of the work artifacts.

These data were later transcribed into an activity tracking log and analyzed by two researchers independently according to the group-discussed coding system. The inconsistencies of their coding were also discussed by the group. We annotate the time to the second whenever peripheral context changed (e.g. an instant message popped up, a project member came to discuss) or the participant performed a switching action (e.g. making a phone call, switching to another application). Related artifacts, tools, and people were recorded. The activity contents like conversation topics and email content were also transcribed. These data combined with participant’s explanations helped us understand the context and underlying relationship of their activities.

3 Fragmented Interleaving Activities

We observed that project manager’s work is very fragmented. They switch between tools, tasks, and people very frequently (Table 1). According to our video-recorded data, they stay on one application for about 2.1 minutes before switching to another; and they could averagely stay on one task thread for about 5.6 minutes before switch; and every 5.2 minutes they would communicate with a different person.

Table 1. Average time distribution in a tool, task, person before switch (seconds)

	Tool	**Task thread**	**People**
Average time before switch	126.87	338.57	311.84

As demonstrated in Figure 1, we selected a typical scenario of a project manager’s work, which illustrated how his work was fragmented. In the only 2.6 hours, his work involved 8 applications, 5 projects (18 task threads), and he communicated with 15 persons.

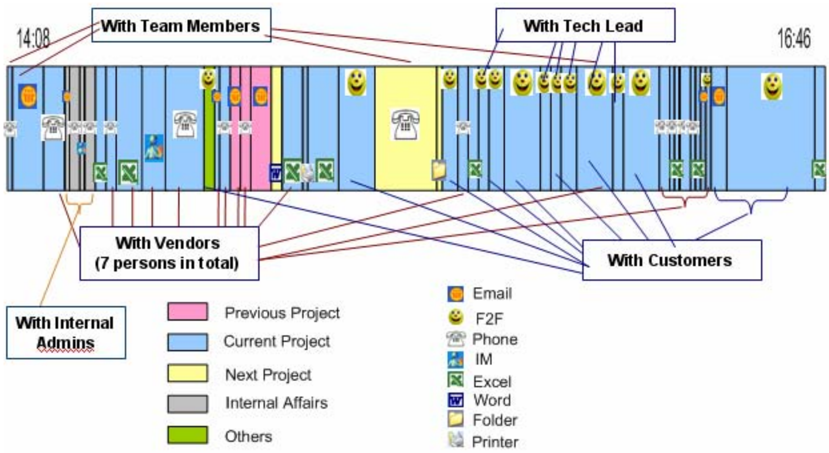


Fig. 1. A typical scenario of a project manager's work

4 Activity Scattered in Tools

Although the project managers were provided integrated platforms (e.g. some project and portfolio management systems) aiming at supporting their collaborative work, they only used conventional office tools like email, instant messaging, and word processor in their daily work. Figure 2 presents the breakdown of time allocated on different activities in the observed project management work. According to the data from our observation, most of a project manager's time (73%) was spent communicating. The major forms of communication were formal meeting, informal discussion, phone, email, and instant message (0.23%, too small to be demonstrated in Figure 2). Actually the other time spent on word processor, power point, and excel was usually communicative because they were reorganizing data from the documents from others and generating new documents to deliver to others (see also Figure 3).

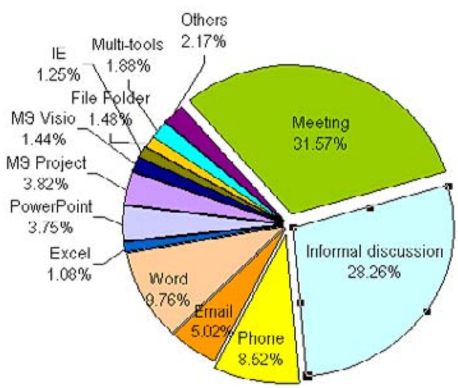


Fig. 2. Time allocation in different applications

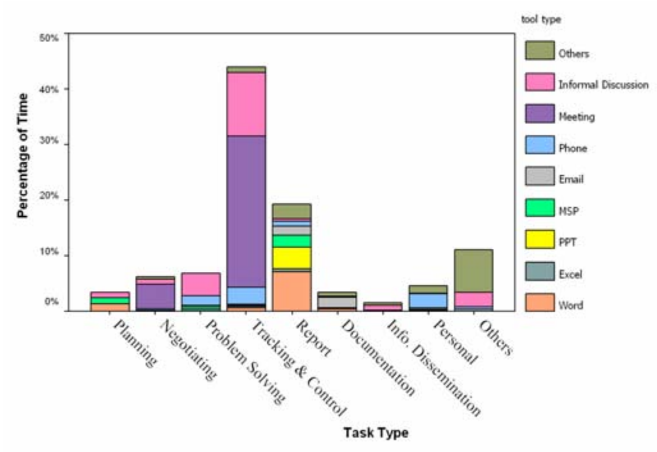


Fig. 3. Time allocation of activities in different applications

We found that project managers communicated similar information through various applications. For instance, if a project manager wanted to get an update on the progress of a member's task, they could send an email to the member; ask them through instant messaging, telephone or if co-located, by walking to their office for a face to face conversation. As a result, the project information was scattered across various systems such as email, reports, meeting minutes, instant messages, and project management tools. Comparing typical collaborative system usages, we found that these systems have their unique applicability, while in many cases those functions overlap with each other (see Table 2). Consequently, project information became not only scattered, which made it hard to locate and track, but also duplicated in different systems, which cost extra effort to do repeated work. While these systems were not integrated with each other, the information delivered was similar or closely related.

Table 2. Comparison of common collaborative systems in project management

	Time/space Matrix	**Other Characters**	**Functions**
Meeting	Synchronous Co-located	Formal, Long	Discussing, Negotiating, Tracking & reporting
Discussion	Synchronous Co-located	Informal Long or short	Discussing, Negotiating, Tracking & reporting, Reminding, Informing
Telephone	Synchronous Distributed	Conversation via voice, Long or short	Discussing, Negotiating, Tracking & reporting, Reminding, Informing, Chatting
Instant Messaging	Synchronous Asynchronous Distributed	Text typing, Short dialogue	Discussing, Negotiating, Tracking & reporting, Reminding, Informing, Chatting
Email	Asynchronous Co-located, Distributed	Can deliver document, Can send to a group at one time, Can be easily tracked	Tracking & reporting, Reminding, Informing, Chatting, Delivering documents, Depositing documents Contextual cueing

5 Activity Organized with Task Threads

Based on the content analysis of the videotaped data, project managers' fragmented work activities were actually interrelated and could be organized by several threads. In Figure 4 a typical day is briefly presented. The project manager originally planned to write a report. During his work, the customer interrupted to ask the project status; his manager asked for another report and checked whether he had sent it out twice; he was also informed to prepare for an unexpected audit, which was urgent and he spared a lot of time for it in the afternoon; there are also weekly meetings that forced him to halt his current work for hours; he also encountered situations like system broken, unfulfilled prerequisites, and missing information that he had to divert from his work. Some threads were completed in single activities, other threads, however, comprised several discrete activities, or even several branch threads, which may also comprised several activities.

In Figure 4 it is presented that most of activities actually have nesting threads. There were related findings such as Gonzalez and Mark's work [7], and they referred these units of work as "working spheres". However, we found these working threads were actually hierarchical. In each activity thread, there could be several digressions and intrusions which initiated certain sub-thread (e.g. check other's progress when writing reports). The working memory and retrospective memory were involved in the active digression and passive intrusion; user usually lost context cues to retrieve memory if they were not properly dealt with.

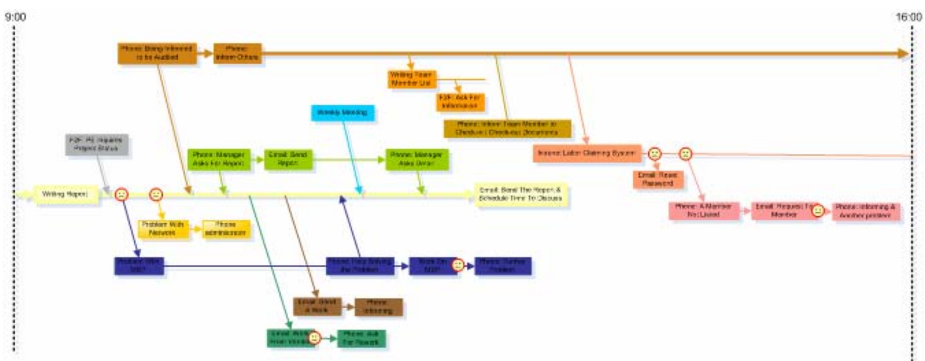


Fig. 4. Activity fragments with underlying threads in a briefed typical day

Actually the terms "thread" and "sub-thread" were all activity threads at different levels. In the high enough level, these threads geared into project tasks as listed in work breakdown structure (WBS). The dozens of activity threads in one day could be categorized into several project tasks (e.g. work on contract, work on testing).

6 Our Model: Reconstructing Collaborative Activity

Project management activities were organized by threads related to WBS, which suggested an opportunity that the seeming disordered work activities could be

organized by aligning with project tasks. As illustrated in Figure 5, the lightweight collaborative activities (activity content) could be integrated and organized in a platform with the threaded list (activity list); and lists of activities could be integrated with and organized by the list of project tasks, which the activities were performed to complete. There have been attempts like the Activity Explorer [19] [21] [22] to integrate various activities in one platform. Further linking the activity list with project task will make the activities much meaningful and manageable.

By aligning with project tasks, the loosely coupled activities with underlying relationships would be tied together and be attached to certain task that served as their goal or scope. On the other hand, the organized activities would absolutely extend project task in smaller granularity, which greatly facilitated project managers to track project information in a rich context as well, as the most of the information was actually conveyed in the lightweight activities.

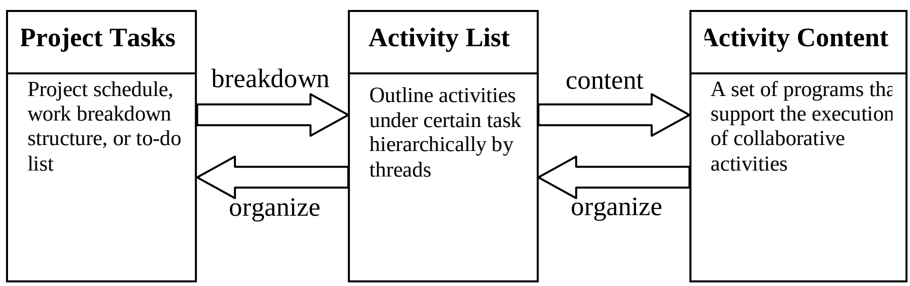


Fig. 5. Organizing lightweight activities with threads by aligning with project tasks

And other information contained in the activities was also captured as a side effect. The time stamp of the activities, for example, was rather critical for tracking project information. Information on the person involved, the activity type as well as the activity topics was provided, which otherwise were usually stored in people's mind and thus hard to retrieve for others.

7 Conclusion

In this paper, we reported an exploratory field studies on lightweight collaborative activity patterns in the scenario of project management. Comparing our study findings with the work which Bannon et al did twenty-three years ago on activity organization, it's surprising to find that complex patterns of interleaved activities performed by users of computer systems didn't change too much with twenty years of HCI and CSCW technology development. People are still using conventional office tools in a loosely coupled way, which makes their work rather fragmented and disordered. And extra effort was required to track, locate and organize the project information which was distributed in different tools and channels, which usually have functional overlap.

Our data indicated that those seeming disordered and fragmented activities actually had underlying threads that geared into project tasks, which explains how project

managers could handle overwhelming information within the mental workload limitation. This also provides implication that the lightweight collaborative activities could be organized by hierarchical threads, by aligning with project tasks.

There have been a lot of studies and findings of managerial work. Our results confirm some of them while have many distinctions. First, we studied collaborative activities in project management scenario. Our observation data were combined with work contents and project contexts. We considered them as cases rather than just statistical data. Second, although our "activity thread" assembles the concepts like "goal-related task" [2], "working sphere" [7], etc., we find these activities threads are actually hierarchical and could gear into project tasks, which suggest a way to organize work activities and project tasks. Our findings indicated that an activity approach could bridge personal ad hoc communication with project or organization collaboration and support the individual in context of organization structures to better manage collaboration.

References

1. Bailey, B.P., Konstan, J.A., Carlis, J.V.: The effects of interruptions on task performance, annoyance, and anxiety in the user interface. In: Hirose, M. (ed.) Human-Computer Interaction INTERACT 2001 Conference Proceedings, pp. 593–601. IOS Press, IFIP (2001)
2. Bannon, L., Cypher, A., Greenspan, S., Monty, M.: Evaluation and analysis of users' activity organization. Transcript of talk delivered at CHI '83 (1983)
3. Bellotti, V., Dalal, B., Good, N., Bobrow, D.G., Ducheneaut, N.: What a to-do: studies of task management towards the design of a personal task list manager. In: The Proceedings of CHI '04., pp. 735–742. ACM Press, New York (2004)
4. Bellotti, V., Ducheneaut, N., Howard, M., Smith, I.: Taking email to task: The design and evaluation of a task management centered email tool. In: The Proceedings of the CHI 2003, pp. 345–352. ACM Press, New York (2003)
5. Card, S.K., Henderson, A.: A multiple virtual - workspace interface to support task switching. In: The Proceedings of CHI '87, pp. 53–59. ACM Press, New York (1987)
6. Czerwinski, M., Horvitz, E.: An investigation of memory for daily computing events. In: The Proceedings of HCI 2002: 16th British HCI Group Annual Conference, pp. 229–245 (2002)
7. González, V., Mark, G.: Constant, constant, multi-tasking craziness: Managing multiple work spheres. In: The Proceedings of CHI '04, pp. 113–120. ACM Press, New York (2004)
8. Harrison, B.L.: An activity-centric approach to context sensitive time management. CHI 2004: Workshop on the Temporal Aspects of Work (2004)
9. Jett, Q.R., George, J.M.: Work interrupted: A closer look at the role of interruptions in organizational life. Acad Mgt. Rev. 28(3), 494–509 (2003)
10. Mark, G., González, V., Harris, J.: No task left behind? Examining the nature of fragmented work. In: The Proceedings of CHI 2005, pp. 321–330. ACM Press, New York (2005)
11. Millen, D.R., Muller, M.J., Geyer, W., Wilcox, E., Brownholtz, B.: Patterns of media use in an activity-centric collaborative environment. In: The Proceedings of CHI 2005, pp. 879–888. ACM Press, New York (2005)
12. Moran, T.P.: Unified activity management: Explicitly representing activity in work-support systems. In: Position paper at ECSCW 2005 workshop: Activity - From a theoretical to a computational construct (2005)

13. Muller, M.J., Geyer, W., Brownholtz, B., Wilcox, E., Millen, D.R.: One-hundred days in an activity centric collaboration environment based on shared objects. In: The Proceedings of CHI 2004, pp. 879–888. ACM Press, New York (2004)
14. Muller, M.J., Wu, M.: How much is shared in a shared activity? In: The Proceedings of the CHI 2006, ACM Press, New York (2006)
15. Oulasvirta, A., Saariluoma, P.: Long-term working memory and interrupting messages in human-computer interaction. Behaviour and Information Technology, pp. 53–64 (2004)
16. Seshadri, S., Shapira, Z.: Managerial allocation of time and effort: the effects of interruptions. Management Science 47, 647–662 (2001)

Part III

Cognitive Modeling and Measuring

Cognitive Model Data Analysis for the Evaluation of Human Computer Interaction

Jeronimo Dzaack[1] and Leon Urbas[2]

[1] Berlin University of Technology, Center of Human-Machine Systems
Franklinstraße 28-29,
Sekretariat FR 2-7/2, D-10587 Berlin
Jeronimo.Dzaack@zmms.tu-berlin.de
[2] Dresden University of Technology, Institute of Automation
D-01062 Dresden
Leon.Urbas@tu-dresden.de

Abstract. In industry and consumer electronic, more and more operative tasks are changing to supervisory control and management tasks. This leads to more complex and dynamic user interfaces (e.g. integrated control interfaces, infotainment systems in cars). Because of the integrated functionality and the complex data structures, these interfaces require more cognitive information processing. Usability of such interfaces can be evaluated by using cognitive modeling to investigate cognitive processes and their underlying structures. So far the explanatory power of cognitive models is limited due to the lack of fine-grained simulation data analysis. Having realized this drawback we developed SimTrA (*Sim*ulation *Tr*ace *A*nalyzer) to simplify the analysis of cognitive models. The tool automatically processes and analyzes data from cognitive models and allows the comparison of simulated data with empirical eye movement data. This paper describes the approach and its implementation. The practicability of SimTrA is demonstrated with an example in the domain of process control.

Keywords: Cognitive Architectures, Eye Movement Data, Analysis, Human Computer Interaction.

1 Introduction

Recent introductions of new information technologies in the field of dynamic human-machine systems (e.g. process control systems in the chemical industry or airplane cockpits) have led to increasing cognitive requirements caused by a shift from manual process operation to the management of complex automated processes. This calls for highly complex dynamic user interfaces. Because of the integrated functionality and the complex data structures, these interfaces require more cognitive information processing. In a practical application cognitive models can be used to evaluate the usability of prototypes. This helps to detect errors in the interaction design of interfaces and gives indications about the cognitive demands of the future user. Cognitive models extend classical usability methods (e.g. questionnaires, observing) and expand the

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 477–486, 2007.

repertoire by cognitive aspects. Consider three groups each developing an interface for a new human-machine system with the same functionality. The three interfaces can differ in appearance, and the choice and arrangement of displayed information. Empirical tests can be conducted to detect the interface that best fits human cognitive demands, or a cognitive model can operate the machine via the interface. The quantitative and normative cognitive data output (i.e. error rates, times, decisions) helps to evaluate the interfaces and to detect the best one for the given purpose. This could shorten the development process and reduce costs and required resources. The potential of cognitive architectures and cognitive models is known in research and development related to usability. However, this method is rarely employed in usability research and development because of a lack of support tools for cognitive modeling. Focus of this research is the analysis of cognitive model data.

2 Cognitive Modeling as Method for the Evaluation of HCI

Cognitive architectures incorporate psychological theories (e.g. visual information processing, decision making) and empirically based representations about aspects of human cognition. There is general agreement that these cognitive aspects are relatively constant over time and relatively task-independent [8]. Therefore cognitive architectures present these aspects in a software framework to explain and predict human behavior in a detailed manner. In this context, a cognitive model can be seen as an application of a cognitive architecture to a specific problem domain with a particular knowledge set. Building a cognitive model, the modeler must describe cognitive mechanisms in a highly-detailed and human-like way. Two levels of cognitive architectures can be differentiated [14]. High-level architectures (e.g. GOMS [4]) describe behavior at a basic level and define interactions as a static sequence of human actions. Low-level architectures (e.g. ACT-R, SOAR or EPIC, see [3] for an overview) describe human behavior at an atomic level. They allow a more detailed insight into cognitive processes than high-level architectures. Most low-level architectures use production systems to simulate human processing and cognition. The use of independent production rules allows cognitive models to react on external stimuli (bottom-up processes) and to model interruption and resumption of cognitive processes in contrast to high-level architectures which are usually controlled top-down. The research presented in this paper uses the cognitive architecture ACT-R [1].

To use cognitive models for the evaluation of Human Computer Interaction, the analysis of the cognitive model data is an important aspect. Most low-level simulation experiments use global information to analyze the model and its fit to empirical data (e.g. errors, times). There are two problems. First, in order to ensure that a cognitive model acts like a human and that this behavior can be explained theoretically, the results with a cognitive model and the human and the internal computations to achieve these results have to be the same [15]. Second, different cognitive models can have the same overall information but can differ in underlying sub-processes. That means that in order to analyze the model's performance on the basis of the integrated psychological theories (e.g. visual perception and processing) it is necessary to observe the global structure and also the microstructure level to identify the underlying processes [6]. The microstructure can be characterized by the sub-processes of the

cognitive model, i.e. repeated short sequences of action such as control-loops. This makes it possible to analyze the kind of computations that lead to a result and to determine the level of correctness of the cognitive model with respect to the underlying psychological theories.

However, using cognitive models for usability evaluation reveals some problems. Cognitive architectures and models are incomplete and do not describe all the processes that are responsible for human cognition [10]. This is due to the partial knowledge of internal cognitive processes and also the fact that it is not necessary to implement cognitive aspects such as esthetics, boredom, fun, or personal preferences in order to explain an effect [3]. Analyzing and comparing cognitive model data with empirical data requires taking these differences into account.

So far, no tools exist for the extraction of fine-grained information from model data for the evaluation of user interfaces.

3 Analyzing Cognitive Model Data

The research project aims to support the analysis of cognitive model data for the evaluation of human computer interaction. For this reason the *Sim*ulation *Tr*ace *A*nalyzer (SimTrA) was developed. The implemented process of SimTrA is divided into three components: (1) preprocessing of the raw data, (2) analysis of the preprocessed data, and (3) comparison with further models or empirical data (see Figure 1). This allows the independent development of each module and the easy extension of its functionality in the future. SimTrA enables the user to apply basal algorithms regarding the process of analyzing and comparing. After each step the processed data is stored in a general-purpose format and can be processed by external tools (e.g. MatLab, R, SPSS). It is possible to import cognitive model data as well as empirical data in the tool. The results are plotted as in classical usability-evaluation methods.

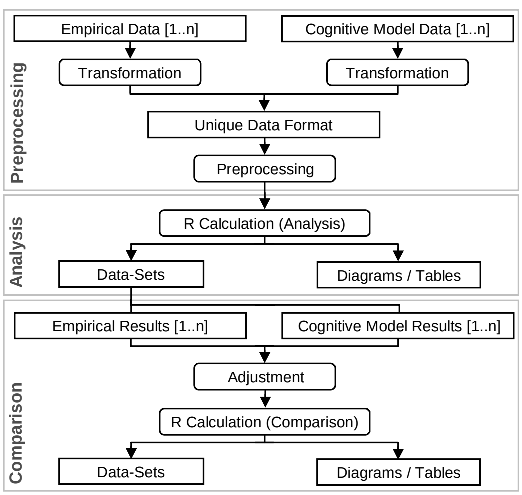


Fig. 1. Overview of the implementation of SimTrA

Eye movement studies in psychological and cognitive research give insights into human behavior and underlying cognitive processes [9, 12, 13]. Cognitive models with ACT-R can process visual information with the integrated visual module, providing spatial and temporal information of the simulated eye movement as found in empirical studies. The extraction of this information provides a way to analyze and compare cognitive model data with empirical data at a global level (e.g. overall number of fixations in a trial and their durations) and a microstructure level (e.g. scanpaths). In a prior study we showed the possibility to extract simulated eye movement data from cognitive models and to compare it with human eye movement data [5]. That is the reason for the implementation of SimTrA for eye movement data.

3.1 Preprocessing

For the preprocessing, raw data of empirical experiments or cognitive model experiments can be imported into SimTrA and transferred into a general-purpose format (see Figure 1) which contains all information on the eye movement. A plug-in has to be integrated into the cognitive model which allows the storage of the current location of focus, timestamp, action and state of the visicon of ACT-R (i.e. state of the simulated interface) for each access of the visual module by the cognitive model (e.g. shift of attention, encoding) in an Extended Markup Language file (XML-File). For the empirical data, the raw data output of the eye movement tracking software is reduced to the current location of focus and timestamp (minimum demands of eye movement studies). The transformation allows processing of data of different quality and origin (e.g. different cognitive architectures, empirical studies) with the same algorithms.

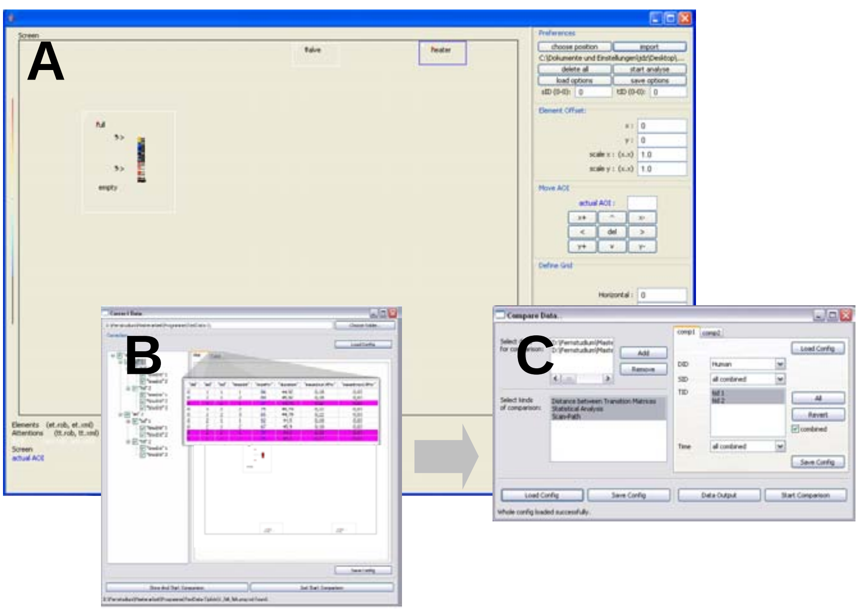


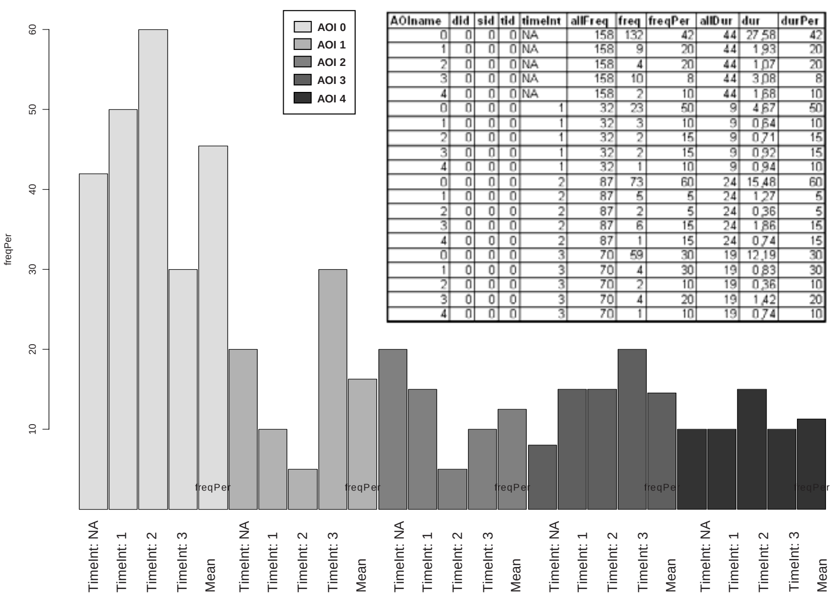
Fig. 2. Screen views of SimTrA. A: Preprocessing. B: Adjustment of data. C: Comparison of data.

While preprocessing each dataset, the stored data is augmented by information on the origin of the data (e.g. cognitive model or empirical data), the subject and the trial identification (i.e. data-Id, subject-Id and trial-Id). Each kind of information can be altered by the user in the preprocessing component of SimTrA.

In this component, the data of different origins can be scaled to a similar resolution by the user to ensure that the data is comparable. To apply eye movement algorithms, the user has to enrich the current data with additional information that is needed for the analysis (i.e. areas of interest, time-intervals) and has to set the output folder. This is done in the preprocessing interface (see Figure 2) by simply clicking in the plotted data-points to set an area of interest (AOI) or by using the tools for the time-intervals and other additional information. The preprocessing of the raw data is finished by choosing the desired analysis methods. All information is stored in XML-files in the output folder, and is available for reuse (e.g. AOIs, time-intervals, and datasets), making it possible to skip the import of the raw data in some subsequent analyses. For this the user can use the *load options* command in the interface.

3.2 Analysis

This component enables the analysis of the preprocessed data (see Figure 1). The algorithms for the analysis are implemented in R, a free tool for statistical computing [11]. The results are saved in tables and as graphical plots (see Figure 3). In this step algorithms are implemented to analyze the transition frequencies, the fixations on AOIs, the spatial density, the statistical dependency in visual scanning, the local scanpaths and a mean for each of these algorithms (for an overview see [13]). In the result



AOIname	dId	sId	tId	timeInt	allFreq	freq	freqPer	allDur	dur	durPer
0	0	0	0	NA	158	132	42	44	27,58	42
1	0	0	0	NA	158	9	20	44	1,93	20
2	0	0	0	NA	158	4	20	44	1,07	20
3	0	0	0	NA	158	10	8	44	3,08	8
4	0	0	0	NA	158	2	10	44	1,68	10
0	0	0	0	1	32	23	50	9	4,67	50
1	0	0	0	1	32	3	10	9	0,64	10
2	0	0	0	1	32	2	15	9	0,71	15
3	0	0	0	1	32	2	15	9	0,92	15
4	0	0	0	1	32	1	10	9	0,94	10
0	0	0	0	2	87	73	60	24	15,48	60
1	0	0	0	2	87	5	5	24	1,27	5
2	0	0	0	2	87	2	5	24	0,36	5
3	0	0	0	2	87	6	15	24	1,86	15
4	0	0	0	2	87	1	15	24	0,74	15
0	0	0	0	3	70	59	30	19	12,19	30
1	0	0	0	3	70	4	30	19	0,83	30
2	0	0	0	3	70	2	10	19	0,36	10
3	0	0	0	3	70	4	20	19	1,42	20
4	0	0	0	3	70	1	10	19	0,74	10

Fig. 3. Example output of SimTrA

tables the data is distinguished by the data-Id, subject-Id and the trial-Id. The absolute and the cumulative number of fixations, fixation durations and general information are stored in a general table. Each algorithm accesses the general table and processes the information needed for the calculation of the respective parameter.

To support the user, plots are generated for each analysis. All input-files (i.e. empirical and cognitive model data), results, general information and the distribution of eye movement data are stored in either tables or plots.

3.3 Comparison

This component enables the comparison of the analyzed data from cognitive models or empirical studies (see Figure 1) and the reapplication of analyses of data files with adjusted data. The comparison allows the rating of the data (e.g. cognitive model vs. empirical data, cognitive model 1 vs. cognitive model 2) with respect to the given context (e.g. empirical findings, psychological theories). The reapplication and the adjustment allows the user to detect missing data and to delete the corresponding dataset and finally to reanalyze the data file.

In both cases the data has to be loaded into the comparison-component. In a first step each data file can be adjusted by the user (e.g. missing data, insufficient data points). In this step all datasets of one data file are displayed and the user can select or deselect the datasets by clicking on them. To support the user, the upper and lower boundary values of the given data for one column of the displayed general table (upper and lower percentile, [2]) are calculated and displayed and a plot of the distribution of the eye movement is shown for each dataset (see Figure 2). The percentile thresholds and the column used for the calculation can be altered by the user. All adjustments are stored in a file in the input folder and can be loaded into the tool if this step is repeated.

In the next step, the type of algorithms and the output folder are chosen by the user and the comparison or analysis is started. The algorithms for the comparison are also implemented in the software R. The comparison is done by descriptive methods because cognitive model data do not show a high variance and statistical interference methods are not applicable in this case. The input, the analyzed data and the compared data are stored in tables and plotted in files.

4 Practical Application: Process Control System

We conducted an study to show the practicability of SimTrA. Therefore, we processed and compared simulated eye movement data of two different cognitive models with empirical data. Eye movements of 12 participants were recorded during their interaction with three complex dynamic interfaces of a process control system microworld (randomized within design, independent variable: version of interface; dependent variables: eye movement data and performance). The task lasted 45 seconds for each interface. The aim was to stabilize the level of liquid in a container, which is moderated by inflow, outflow and evaporation, between two boundary values. The parameters can be regulated by a valve and a heater connected with the container. The three interfaces differ in the position of the interaction and feedback elements of the interface (see Figure 4).

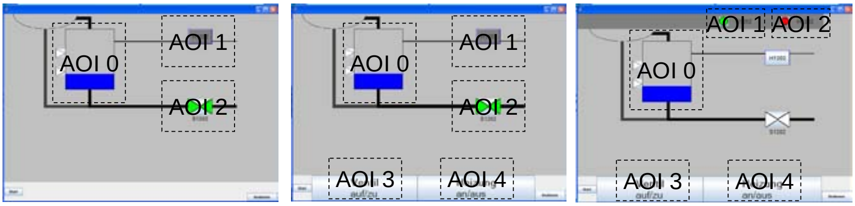


Fig. 4. Three interfaces for one task. A: Combined information and interaction elements. B and C: Distinct information and interaction elements.

For the model based investigation we developed two ACT-R 6.0 models representing the internal knowledge of an experienced user familiar with the technical process (as found in a survey of experts). One model is implemented in the minimal control principle style suggested by Taatgen [16]. That allows the cognitive model to react to external stimuli (bottom-up control). The other model is implemented with a fixed sequence of actions moderated by states (top-down control).

In both the empirical experiment and the simulation, x- and y-coordinates of gaze targets with timestamp and the overall performance of task were recorded. SimTrA was used to analyze the microstructure behavior and to compare the empirical data with the cognitive model data.

4.1 Model Data and Empirical Data

For the evaluation of the practicability of SimTrA, we compared the interaction of the humans (H), cognitive model 1 (bottom-up control, CM 1) and cognitive model 2 (top-down control, CM 2) with the three different interfaces (3 x 3 design). We investigated the overall data and microstructure data.

For overall data analysis, we determined the percentage of the time where the level is between the two boundary values (i.e. in control) and the amount of interactions. The performance of the three groups for all 3 interfaces is nearly the same (H: 90 % in control / Heater: 1 press / Valve: 8 presses, CM 1 & CM 2: 85 % in control / Heater: 1 press / Valve: 9 presses).

For the microstructure analysis we defined three or five AOIs (see Figure 4) and applied all integrated algorithms of SimTrA. Because of the large numbers of tables and plots we discuss an extract of the analyzed data. Ten datasets could be used for the analysis after the adjustment. Table 1 shows the data for the comparison of the local scanpaths (consistent patterns of consecutive fixations) of actual perceptions (closed-loop control) [7]. All theoretical triple of AOIs are determined whereby any sequence of fixations falling into the same AOI is treated as a single gaze fixation and the number of occurrence in the whole sequence of AOIs is assigned to them (e.g. AOIs: 1, 2 - Sequence: 2122112 - triple: 121: 1, 212: 2). Ordering these by frequency shows the most important local scanpaths. The local scanpaths make it possible to determine important visual scanpaths and help to detect faults and obstacles of the interface design that can influence the information encoding and processing [13].

Table 1. Comparison of the frequencies of local scanpaths greater 3 % as found in the empirical study in percent for the three interfaces (H: Human, CM 1: bottom-up model, CM 2: top-down model)

Interface 1				**Interface 2**				**Interface 3**			
Path	**H**	**CM 1**	**CM 2**	**Path**	**H**	**CM 1**	**CM 2**	**Path**	**H**	**CM 1**	**CM 2**
0>2>0	26.1	36.4	9.9	0>2>0	11.6	3.0	37.8	0>3>0	14.6	11.9	9.5
2>0>2	21.7	33.3		0>3>0	11.6	24.2	9.5	0>1>0	6.4	4.8	
1>0>2	8.1	9.1	9.9	0>1>0	7.5	15.2		2>0>3	6.4	2.4	
2>0>1	8.1		9.9	2>0>2	5.5		29.7	3>0>1	6.4	4.8	
0>1>0	6.8	3.0	39.4	3>0>1	5.5	12.1		3>0>3	5.3	7.1	
2>1>0	6.8	9.1	1.4	1>0>3	4.8	12.1		0>4>0	4.7	2.4	1.4
				3>0>2	4.8		8.1	0>1>2	4.1		
				2>0>3	3.4	3.0	9.5	1>2>0	4.1	2,4	1.4
								1>0>1	3.5		
								3>0>2	3.5		8.1
Sum	**83.8**	**90.9**	**70.4**	**Sum**	**54.8**	**69.7**	**94.6**	**Sum**	**59.1**	**35.7**	**20.3**

4.2 Discussion

The empirical study and the simulation show that the three groups do not differ essentially in the overall behavior in the interaction with the three interfaces. Table 1 shows the comparison of the local scanpaths of the three groups (i.e. human, bottom-up model and top-down-model). Scanpaths with a frequency < 3% were excluded and the remaining important scanpaths were used for the comparisons. The table shows that the frequencies and the rank order of the predicted local scanpaths is different between all groups. Nevertheless the bottom-up model (CM 1) seems to have a better fit to the empirical data than the top-down model (CM 2) based on the rank order and the reached cumulative percents of predicted local scanpaths. This allows the conclusion that the cognitive processes that moderate human eye movement and the simulated cognitive processes are different. We infer that the cognitive models and its architecture need to be improved in several respects (e.g. update circles and uncontrolled eye movements). Taking into account all processed data and the extracts shown in this paper, we conclude that SimTrA is able to extract and process cognitive model as well as human eye movement data. SimTrA helps to detect differences and similarities between different cognitive models for the same task as well as between simulated and human eye movement data.

5 Conclusion

We have developed SimTrA to analyze cognitive model data at a microstructure level. In this paper we show the internal approaches and the practicability of SimTrA for the analysis of simulated and human eye movement data. Cognitive models allow a more detailed investigation of simulated cognitive processes (e.g. decision-making, arousal)

than the eye movement data. But it has not been established which cognitive processes are executed during human information processing. It is therefore better to rely on observable parameters than on uncertain internal processes.

SimTrA enables the detection of diverse microstructure behavior between different cognitive models that seem to behave equally from an overall point of view. There seem to be three possible areas for the application of SimTrA. First, it allows the verification of distinctive psychological theories integrated in cognitive models. The analysis and comparison of cognitive models implemented with different theoretical groundings can help to identify varieties and similarities in these theories. Second, the analysis and comparison of empirical and simulated data helps to improve the cognitive model and its internal structure. Part of this research is the improvement of cognitive models. Our results accentuate the need to adjust cognitive models at a microstructure level to make human behavior and cognitive models more comparable. And finally, the cognitive model data analysis for the evaluation of human computer interaction is enabled. Altogether, this leads to more detailed model data and allows a better prediction of human behavior for usability questions.

Further work is needed on the extension of SimTrA. It is necessary to evaluate which additional analysis and comparison algorithms can be integrated.

To summarize, our tool enables the analysis and the comparison of cognitive model behavior with human behavior at a microstructure level. But the cognitive model data analysis must be enhanced by a greater variety of algorithms. This could lead to an increased application of cognitive models in the usability evaluation of computer interfaces.

Acknowledgment. This project is funded by Deutsche Forschungsgemeinschaft (Research Training Group 1013 prometei).

References

1. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. Psychological Review 111, 1036–1060 (2004)
2. Bortz, J.: Statistik für Sozialwissenschaftler. Springer-Verlag, Heidelberg (1999)
3. Byrne, M.D.: Cognitive Architectures. In: Jacko, J., Sears, A. (eds.) Handbook of Human-Computer Interaction, pp. 97–117. Lawrence Erlbaum Associates, Hillsdale, NJ (2003)
4. Card, S.K., Moran, T.P., Newell, A.: The Psychology of Human-Computer Interaction. Lawrence Erlbaum Associates, Hillsdale, NJ (1983)
5. Dzaack, J., Urbas, L.: Kognitive Modelle zur Evaluation der Gebrauchstauglichkeit von Mensch-Maschine Systemen. In: Grandt, M., Bauch, A. (eds.) Cognitive Systems Engineering in der Fahrzeug- und Prozessführung. Deutsche Gesellschaft für Luft- und Raumfahrt, Bonn, pp. 305–318 (2006)
6. Gray, W.D., Sims, C.R., Schoelles, M.J.: Musings on Models of Integrated Cognition: What Are They? What Do They Tell Us that Simpler Approaches Cannot? How Do We Evaluate Them? In: Proceedings of the 7th ICCM, pp. 12–13. Lawrence Erlbaum, Hillsdale, NJ (2006)
7. Groner, R., Walder, F., Groner, M.: Looking at faces: local and global aspects of scanpaths. In: Gale, A.G., Johnson, F. (eds.) Theoretical and Applied Aspects of Eye Movement Research, pp. 523–533. Elsevier, Amsterdam (1984)

8. Howes, A., Young, R.M.: The role of cognitive architecture in modeling the user: Soar's learning mechanism. Human-Computer Interaction 12(4), 311–343 (1997)
9. Just, M.A., Carpenter, P.A.: Using eye fixations to study reading comprehension: Individual differences in working memory. Psychological Review 99, 122–149 (1984)
10. Opwis, K.: Kognitive Modellierung: Zur Verwendung wissensbasierter Systeme in der psychologischen Theoriebildung. Huber, Bern Göttingen Toronto Seattle (1992)
11. R Development Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2006)
12. Rayner, K.: Do eye movements reflect higher order processes in reading? In: Groner, R., d'Ydewalle, G., Parham, R. (eds.) From Eye to Mind: Information Acquisition in Perception, Search, and Reading, Elsevier, New York (1999)
13. Rötting, M.: Parametersystematik der Augen- und Blickbewegungen für arbeitswissenschaftliche Untersuchungen. Shaker, Aachen (2001)
14. Salvucci, D.D., Lee, F.J.: Simple Cognitive Modeling in a Complex Cognitive Architecture. In: Proceedings of the CHI '03, pp. 265–272. ACM Press, New York (2003)
15. Strube, G.: Cognitive Modeling: Research Logic in Cognitive Science. In: Smelser, N.J., Baltes, P.B. (eds.) International encyclopedia of the social and behavioral sciences, pp. 2124–2128. Elsevier, Oxford (2001)
16. Taatgen, N.A.: The Minimal Control Principle. In: Gray, W. (ed.) Integrated Models of Cognitive Systems. Oxford University Press, Oxford (in press)

Automatic Detection of Interaction Vulnerabilities in an Executable Specification

Michael Feary

NASA Ames Research Center
MS 262-4
Moffett Field, CA 94035, USA
michael.s.feary@nasa.gov

Abstract. This paper presents an approach to providing designers with the means to detect Human-Computer Interaction (HCI) vulnerabilities without requiring extensive HCI expertise. The goal of the approach is to provide timely, useful analysis results early in the design process, when modifications are less expensive. The twin challenges of providing timely and useful analysis results led to the development and evaluation of computational analyses, integrated into a software prototyping toolset. The toolset, referred to as the Automation Design and Evaluation Prototyping Toolset (ADEPT) was constructed to enable the rapid development of an executable specification for automation behavior and user interaction. The term executable specification refers to the concept of a testable prototype whose purpose is to support development of a more accurate and complete requirements specification.

Keywords: automation design, automation surprise analysis.

1 Introduction

This paper presents an approach to providing designers with the means to detect Human-Computer Interaction (HCI) vulnerabilities without requiring extensive HCI expertise. The goal of the approach is to provide timely, useful analysis results early in the design process, when modifications are less expensive. The approach consists of the development of computational Human-Automation Interaction (HAI) vulnerability analyses, integrated into a software prototyping toolset. The toolset, referred to as the Automation Design and Evaluation Prototyping Toolset (ADEPT) was constructed to enable the rapid development of an executable specification for automation behavior and user interaction. The term executable specification refers to the concept of a testable prototype whose purpose is to support development of a more accurate and complete requirements specification.

The paper specifically focuses on the evaluation of the HAI vulnerability analyses' operational validity. Operational validity refers to the effectiveness of the analyses in predicting actual difficulties in operation. The evaluation was performed by comparing vulnerability predictions from the tool to performance of participants operating two device prototypes created in ADEPT.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 487–496, 2007.

The metrics for determining whether or not an identified automation vulnerability causes difficulty in performance were defined as (1) failures in completing realistic tasks within a given amount of time, and (2) difficulty in predicting future automation behavior. The vulnerability analyses were designed to identify vulnerabilities known to cause difficulty in operation [1,2,3,4].

The vulnerability analyses include:

- Moded Inputs
- Armed Behaviors
- Automatic Behaviors
- Inhibited Inputs
- Similar Feedback

Moded Inputs Analysis

Moded Inputs are user interface objects (e.g., buttons, knobs, etc.) that may result in different device behaviors when an action is taken upon them (e.g., pressed, rotated, etc.). Moded Inputs may make it difficult to predict future automation behavior. A moded input is defined formally in ADEPT as a 'user action which, depending on the mode of the device, can lead to more than one device behavior'.

An example of a Moded Input in the study is a multi-function remote control that can control a television or a video recorder. Depending on the "mode" of the remote control, the power button on the remote control will turn the television or recorder on or off.

The vulnerability associated with moded inputs is most commonly referred to as "mode confusion" [3,5,6,7]. This difficulty is compounded if sufficient feedback is not provided.

Armed Behavior Analysis

Armed Behaviors are device behaviors that require more than just a user interface action for engagement. Armed behaviors are particularly troublesome because a delay usually exists between the user interface action and the time at which all of the conditions for engagement are satisfied.

An example from aviation is the Korean Airlines 007. Although the true cause of this accident may never be known, a probable explanation for this accident is that the crew had "armed" the inertial navigation system, but it did not meet all the necessary conditions for engagement. The crew did not notice that the aircraft was proceeding off-course when they transgressed Russian airspace and were shot down [8].

Automatic Behavior Analysis

Automatic Behaviors are device behaviors that engage independent of user input. Automatic Behaviors are similar to armed behaviors from a HAI standpoint. The difference between the two is that armed behaviors require a user interface action to initiate the automation behavior, while automatic behaviors do not. When coupled

with inadequate feedback, these behaviors have been referred to as “strong and silent” [4].

An example of an automatic behavior is airspeed envelope protection in certain aircraft that automatically engages and increases engine thrust if the airspeed drops below a specific airspeed threshold.

Inhibited Inputs Analysis

An *Inhibited Input* vulnerability is a user interface input action that does not result in a behavior change. Similar to Moded Inputs, Inhibited Inputs make it difficult to predict what the device will do after a user action. Although the vulnerabilities related to inhibited inputs are indirect, they greatly affect the user’s understanding of automation behavior. As such, it is difficult to find incident and accident reports for which inhibited inputs are a contributing factor, however it has been documented in controlled studies that inhibited inputs create confusion.

Examples of inhibited inputs in the video recorder example used in the study, occur while the recorder is in the record mode, which disables all buttons except for the stop and power buttons.

Vakil (1998, 2000) and Javaux (1998) have identified an inhibited input vulnerability as the source of the confusion that can occur when a flight crew attempts to engage a new mode while Approach Mode is active on some modern commercial aircraft.

Similar Feedback Analysis

A *Similar Feedback* vulnerability is present when the same interface objects are used to display the information content for more than one device behavior. In addition to being identified as a vulnerability [9,10]. Similar Feedback vulnerabilities compound other automation surprise vulnerabilities. Feary et. al (1998) examined experienced pilots’ knowledge of aircraft automation behaviors, and demonstrated a lack of pilot knowledge resulting from similar annunciations. The pilots in that study also showed significant improvements in automation behavior prediction when they were given feedback that matched the autopilot behavior.

In the aircraft involved in the Korean Airlines 007 accident, the display of the inertial navigation mode was the same whether the system was “armed” or “engaged”. This made it difficult for the pilots to determine the actual automation behavior.

2 Method

The operational validity evaluation was conducted on two device prototypes. The first device—a video recorder remote control (Figure 1)—was chosen as an example of a “walk-up and use” device (i.e., a person familiar with the goals for using the device should not require special training).

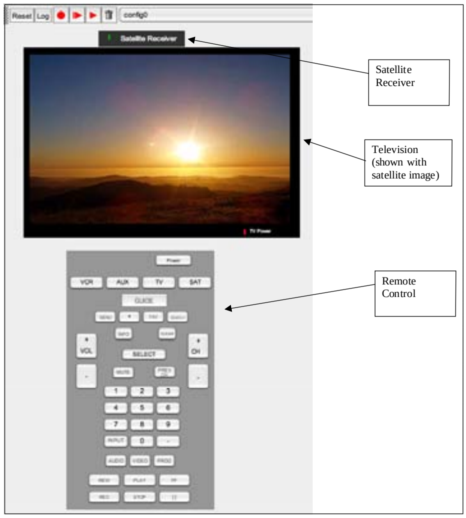


Fig. 1. Remote Control Prototype

The second device—an autopilot for a modern transport category aircraft (Figure 2)—was chosen as an example of a more complex device that does require specific training. The autopilot was chosen to present a challenge to the capabilities of the analyses: the design and training requirements for autopilots are very stringent due the safety critical nature of the commercial aviation.

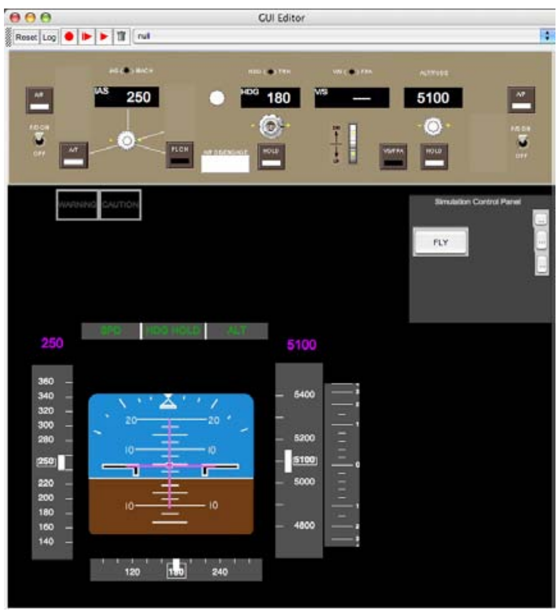


Fig. 2. The autopilot prototype

Participants

Eighteen pilots participated in the evaluation. The pilots ranged in age from 20–46, and each held at least a current private pilot's license. All pilots had a minimum of 200 hours of flight experience. The participants were also screened so as to have no formal Human-Automation Interaction training or expertise.

Procedure

The evaluation was conducted by analyzing the two devices in ADEPT, running domain expert participants through a representative series of tasks on the two device prototypes, and then comparing the results. The evaluation used two measures to evaluate the analyses' prediction ability: task performance, and automation behavior prediction performance.

An accurate prediction of task performance is the ultimate goal of the development of the automation vulnerability analyses, however it would be very optimistic to think that the automation surprise vulnerability analyses alone would be accurate predictors of task completion performance. There are many variables that affect the ability of a user to complete a task and only a small number of those variables are considered in the analyses.

To focus the evaluation of the analyses on automation surprise vulnerabilities, an additional measure referred to as automation behavior prediction performance was also evaluated. Performance on this measure was obtained by asking the participants to select the next automation behavior from a list of potential future behaviors.

Data from the tasks was collected in two ways. First, the experimenter collected task performance data on an evaluation form. The same experimenter (the author) collected data for all 15 participants. A more detailed backup of the data was collected through a data logging facility built into the prototype. Before each task, the experimenter activated the data logging facility, which kept a record of the participants' actions, automation behaviors and state values, and times.

3 Results

The results of the evaluation showed that the analyses predictions did not show significant correlation to the users' performance on the tasks, the predictions were a significantly correlated to the participants' ability to determine future automation behavior.. As described earlier this result can be explained by the influence of additional complexity in predictions of task performance, including the perceptual quality of the interface feedback, the familiarity with the task, and the ability to use heuristics (i.e. process of elimination).

Task Completion Analysis Results

Task completion performance was determined by the ability of the participant to complete the each task within thirty seconds, although none of the participants failed any task because to time limit constraints. The tasks were ranked in order of predicted difficulty, defined as the number if identified vulnerabilities present. Figure 3 shows the results of the participants' task completion performance compared with the task difficulty predicted by the analyses.

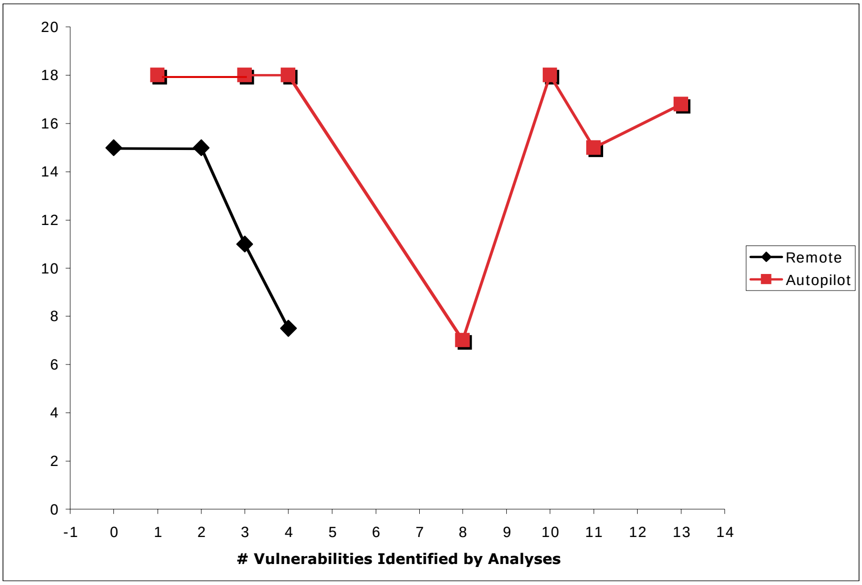


Fig. 3. Analysis Task Completion Results (note that two tasks contained 10 vulnerabilities, and five tasks contained 13 vulnerabilities)

A spearman rank correlation coefficient was computed and showed that the predictions for the remote control tasks were significantly correlated with user performance (r_s=0.97, p<.02, n=15). Although the analyses' predictions for the autopilot tasks were not as accurate, the results show a weak trend of prediction versus user performance (r_s=0.53, p<.1, n=18).

Automation Behavior Prediction Analysis Results

There were only two remote control behavior prediction questions. Both of these were related to which device the (video recorder or television) would be affected by a button press on the remote control.

The participants were asked six autopilot behavior prediction questions. Two of the six autopilot automation behavior prediction questions asked were not used for the evaluation. The questions, with only one vulnerability each, were not counted in the analysis for two reasons: the experiment prototype was determined to not have enough fidelity for the participants to reasonably be expected to answer the questions correctly; and both questions were related to airspeed envelope protection. The automation behavior questions evaluation required that the simulation be paused immediately before the engagement of the automation behavior, and given the rapid onset of airspeed envelope protection, it was not always possible for the experimenter to pause the simulation before envelope protection engaged.

Three of the remaining four questions required responses in terms of the future behavior of the autopilot. The fourth question (three vulnerabilities) asked the participant to predict a mode change, which involved computing an altitude value.

This value was judged to be answered correctly if the participant responded within 100 feet of the correct value.

Similar to the results of the remote control, the analyses were an accurate predictor of which automation behavior prediction questions participants would have difficulty with. Figure 4 shows the results of the participants' performance in response to the automation behavior prediction questions.

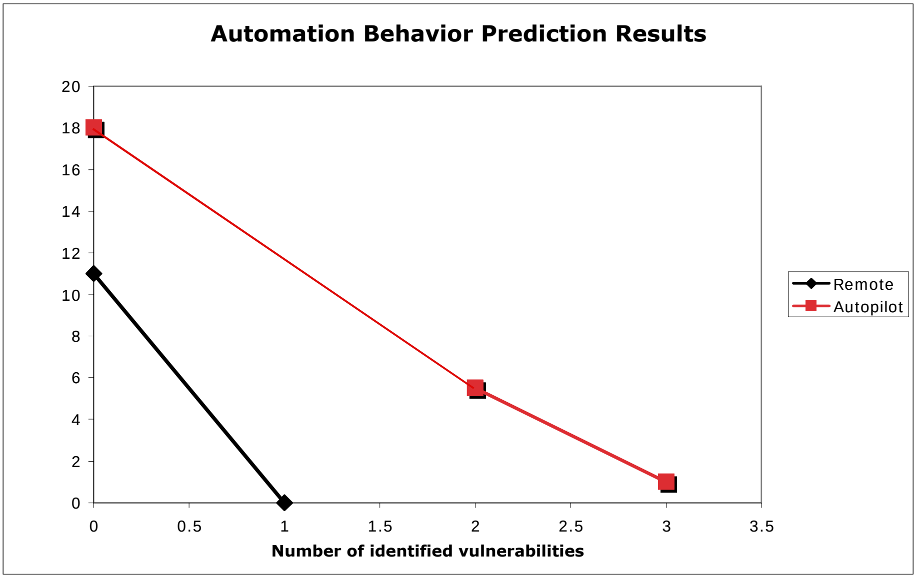


Fig. 4. Automation behavior prediction results (note that the two autopilot questions with two vulnerabilities each are marked by only one point)

Figure 4 shows that the analyses accurately predicted the difficulty of the remote control behavior questions. A spearman rank correlation coefficient was computed for the autopilot behavior questions and showed that the predictions were also correlated with user performance. (r_s=0.95, p<.05, n=18).

4 Discussion

Although the automation vulnerabilities identified by the analyses were poor predictors of task difficulty, the analyses were accurate predictors of the difficulty the participants faced when answering the automation behavior prediction questions. Like the results from the remote control evaluation, this is particularly encouraging, as this is the measure that focuses most upon potential Human-Automation Interaction difficulties.

It is also possible that the analyses' prediction of task difficulty may be better than it initially appears. A problem that became apparent during the experiment is that the button on the Lateral Target Selector knob is easily recognized when implemented in

a 3-dimensional hardware, but difficult to distinguish in the 2 dimensional software representation. As a result, even though it was described in the training, the participants incurred many extra actions looking for the lateral target selector knob on the first task.

There are a number of possible explanations for the analyses task performance results.

The most likely explanation for the inaccuracy of the analyses predictions is that the means of rating the vulnerabilities is immature. The analyses are not currently intended to provide ratings, however, ratings are needed evaluate the quality of the identified vulnerabilities The analyses are intended to be useful for detailed design work, and to identify areas that should be examined further. The results suggest that single vulnerabilities appear to be good predictors of difficulty versus no vulnerabilities, but are currently less useful for comparing two tasks with multiple vulnerabilities.

Closely related to this point, is the absence of weighting of the different vulnerabilities. It is extremely unlikely that all vulnerabilities would have the same effect on user performance. Prior to the study various methods were investigated for weighting the ratings of the vulnerabilities, however no empirical method for assessing the impact of the different vulnerabilities was identified. During the study it was observed that some vulnerabilities appeared to have a greater impact on task performance than others. Additionally, the results of the remote control study suggest the effect that differentiating the vulnerabilities could have improved the performance of the analyses predictions.

An example of the possible effects of weighting was alluded to earlier when describing the “Armed Behavior” vulnerability. The HAI vulnerability is affected by the length of time between user action and engagement (i.e. if the armed behavior engages immediately after the user action, there is not much of an HAI vulnerability). However, as described earlier, the analysis only identified the existence of an armed behavior, not the amount of time beween arming and angagement.

This may provide an explanation for the inaccuracy in the prediction the anomalous autopilot task (shown as the fourth autopilot task from the left in figure 4). This was an off-nominal task, and it is possible that off-nominal tasks and the associated behaviors may cause more difficulty than some of the other vulnerabilities. The remote control results suggest that the weighting may be important, as the analyses provided better results when identifying singular vulnerabilities, however further testing is needed.

Second, the analyses do not evaluate the “look and feel” of the interface. This absence is intentional, as the focus of these analyses is to examine what types of analyses can be formalized, and added as a supplement to traditional interface analysis, but has been proven to be a good indicator of usability difficulties. For example, the affordances work [13] and label following work [14] indicates that participants are likely to select interface objects which have labels which closely match the task description, even if they have been trained in the functions of the objects.

Third, the analyses do not account for sequence dependency or task context. The different patterns in data for the analysis of the automation behavior prediction questions and data for the analysis of the task steps seem to support this. This was

accounted for in the vulnerability point scoring by scoring all of the possible vulnerabilities regardless of sequence, however, the accuracy of the analysis predictions would likely increase if a more systematic method of analyzing the different possible sequences.

Fourth, the results for the prediction of automation behavior tasks suggest that the lack of evaluation of the monitoring behaviors may have impacted accuracy of the task predictions. The task decomposition accounts for monitoring behaviors and the vulnerabilities associated with the monitoring behaviors, however the experiments did not evaluate the effects of the vulnerabilities attached to the monitoring behaviors for the operational tasks evaluated. In contrast, the vulnerabilities associated with the monitoring behaviors could be the focus of an automation behavior prediction question, as illustrated by the first question in the autopilot evaluation which asked the participant to predict the when an armed behavior would engage.

Fifth, the autopilot results may have also been disrupted/exaggerated by a lack of aural and vestibular feedback. Participants who initially made in incorrect action but corrected the action before completing the task were scored as completing the task. A good example of this is the anomalous autopilot task shown in figure 4 (fourth task from the left). It is possible that with additional aural and vestibular feedback, the participants would have corrected their actions. Similar to the fourth point, the results of the automation behavior prediction questions support this, as the directed questions focus on automation knowledge rather than feedback, and monitoring skills.

Sixth, unlike the remote control, the autopilot is a device that is expected to require some amount of training to use. Training is provided specifically to mitigate the errors caused by a complex environment, a complex device, and/or a complex interface. Since the analyses make predictions based on the complexity of the device and interface, training may mitigate the impact of the predictions, and quality or comprehensiveness of the training may lead to differences in performance on certain tasks. For example a majority of training using autopilots for transport category aircraft is spent developing skills for responding to emergency or abnormal situations. As such the certain functions of the autopilot will receive more practice than others [11,15,16].

Seventh, it is possible multiple vulnerabilities may interact with each other to impact the predictions. For example, the autopilot contained inhibited behaviors, which, may impact the understanding of the user, and affect the way the user interacts with the device [11,15,16]. In fact the only autopilot input used by the participants that was not inhibited at some time during the evaluation was the Vertical Speed button, which would engage whenever the participant pressed the button. However it would not always engage the participant’s desired mode. Again, the automation behavior prediction question data supports this by focusing on automation understanding, whereas the educated guesses by the participant may obscure the level of automation understanding when using task data alone.

Although the analyses predictions do not appear to accurately predict the difficulty of tasks for the autopilot for the reasons discussed above, the analyses did appear to accurately predict the difficulty of questions about the automation, and this is a significant finding towards the formalization of usability metrics from an automation behavior based perspective, compared to existing interface based usability techniques.

References

1. Billings, C.: Human-Centered Aviation Automation: Principles and Guidelines. In: NASA Technical Memorandum 110381, NASA Ames Research Center, Moffett Field, CA, USA (1996)
2. Bureau of Air Safety Investigation, Advanced Technology Aircraft Safety Survey Report, Bureau of Air Safety Investigation, Civic Square ACT, Australia (1998)
3. Federal Aviation Administration, The interface between flightcrews and modern flight deck system. FAA human factors team. Federal Aviation Administration. Washington, DC (1996), Available at http://www.faa.gov/avr/afs/interface.pdf
4. Sarter, N., Woods, D.D.: Strong, Silent, and 'Out of the Loop'. CSEL Report 95-TR-01, Columbus, OH, USA (February 1995)
5. Degani, A., Shafto, M., Kirlik, A.: Modes in Human-Machine Systems: Review, Classification, and Application. International Journal of Aviation Psychology 9(2), 125–138 (1999)
6. Sherry, L., Feary, M., Polson, P., Mumaw, R., Palmer, E.: A Cognitive Engineering Analysis of the Vertical Navigation (VNAV) Function. In: NASA Technical Memorandum 2001-210915, NASA Ames Research Center, Moffett Field, CA, USA (2001)
7. Thimbleby, H.: User interface Design. Addison-Wesley, Wokingham, England (1990)
8. Degani, A.: Taming HAL: Designing Interfaces Beyond 2001. Palgrave MacMillan, Hampshire, England (2004)
9. Vakil, S., Hansman, R.J.: Predictability as a Metric of Automation Complexity, In: Human Factors & Ergonomics Society 41st Annual Meeting, pp. 70–74 (September 1997)
10. Vakil, S.: Analysis of Complexity Evolution Management and Human Performance Issues in Commercial Aircraft Automation Systems. PhD Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA. (2000)
11. Javaux, D.: Explaining Sarter and Woods' Classical Results. The Cognitive Complexity of Pilot-Autopilot Interaction on the Boeing 737-EFIS, In: the Proceedings of the 3rd Workshop on Human Error, Safety, and System Development (HESSD'98), Liege, Belgium, pp. 62–77 (June 7-8, 1998). Poller, M.F., Garter, S.K.: The effects of modes on text editing by experienced editor users, Human Factors, 26(4), 449–462 (1984)
12. Feary, M., McCrobie, D., Alkin, M., Sherry, L., Polson, P., Palmer, E., McQuinn, N.: Aiding Vertical Guidance Understanding. In: NASA Technical Memorandum NASA/TM-1998-112217, Ames Research Center, Moffett Field, CA (1998)
13. Norman, D.A.: Cognitive Engineering. In: User Centered System Design. Norman, D.A., Draper, S.W. (eds.) Lawrence Erlbaum Associates, Hillsdale, NJ, USA, pp. 31–61 (1986) Originally: Steps Towards a Cognitive Engineering. Technical Report, Program in Cognitive Science, University of California, San Diego (1981)
14. Polson, P.G., Lewis, C.H.: Theory-based design for easily learned interfaces. Human-Computer Interaction 5, 191–220 (1990)
15. Feary, M., Sherry, L., Polson, P., Fennel, K.: Incorporating Cognitive Usability into Software Design Processes. In: Harris, D., Duffy, V., Smith, M., Stephandis, C. (eds.) Human-Centered Computing: Cognitive, Social, and Ergonomic Aspects, vol. 3, pp. 427–431. Lawrence Erlbaum, Mahweh, NJ (2003)
16. Sherry, L., Fennell, K., Feary, M., Polson, P.: A Human-Computer Interaction Analysis of Flight Management System Messages, NASA/TM-2005-213459. Ames Research Center, Moffett Field, CA, USA (2005)

ATC CTA: Cognitive Task Analysis of Future Air Traffic Control Concepts

Brian Hilburn*

CHPR, The Hague, The Netherlands
hilburn@chpr.nl

Abstract. Traffic Collision Avoidance System (TCAS) is a flight deck tool to display and help avoid proximate air traffic. Until now, TCAS information has not been presented to the air traffic controller. Attention has focused in recent years on the potential benefits of "downlinking" to the air traffic controller TCAS Resolution Advisories (RAs) in near real time, Such presentations, it is thought, could benefit situation awareness and joint decision making between controller and pilot.

A cognitive task analysis (CTA) was recently conducted into the present-day and future RA Downlink (RAD) operational concepts. The aim was to identify the cognitive elements underlying performance in the RAD scenarios, and to hopefully identify potential error mechanisms. On the basis of a functional task description and cognitive walkthroughs, CTA proceeded to decompose the tasks imposed on the controller. The impact of various specific non-nominal events (e.g. pilot reports RA, but does not initiate an evasive maneuver) was investigated. Finally, a set of cognitive elements and potential error mechanisms was identified.

1 Background

Traffic Collision Avoidance System (TCAS) is a 'last-resort' method of preventing mid-air collisions or near-collisions between aircraft. TCAS produces vertical collision avoidance advice in Resolution Advisory (RA) messages and displays these to the flight crew roughly 15 to 35 seconds before closest point of approach. A 2002 midair collision over Überlingen Germany was triggered, in part, by incompatible clearances issued by air traffic control (ATC) on the one hand, and the airborne TCAS on the other. Until now, ATC has not had access to TCAS indications as they are displayed to the pilot. Instead, the controller currently relies on the flight crew to inform them of any deviation from clearance due to an RA and when the aircraft is clear of conflict. If this information is delayed or not received, the controller is unaware and may therefore attempt to resolve the conflict by issuing (potentially incompatible) instructions to the incident aircraft, thereby introducing the risk that the pilot may choose to override TCAS advisories. It was just this situation that occurred over Überlingen.

Increased attention is now being paid to the possible benefits of displaying to ATC any RA messages as they are displayed to the pilot. Such alerting, it is thought, could

* The views presented here are those of the author, and do not necessarily reflect those of any agency.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 497–503, 2007.

benefit situation awareness and joint decision making between controller and pilot. Under EUROCONTROL's Feasibility of ACAS Downlink Scenario (FARADS) program, an operational concept is being defined in which cockpit-generated RAs would trigger a downlink to ATC details of the conflict, including involved aircraft, prescribed evasive maneuver and clear-of-conflict indications. As part of this effort, a Cognitive Task Analysis (CTA) was conducted to help compare the potential for human error under current RA operations and the RAD operational concept.

Task analysis refers to a family of techniques used to describe and analyze operator performance within a human-machine system [1]. In all, it seems that at least three dozen major task analysis techniques have been used over the years [2; 3]. All task analysis techniques aim to decompose complex system tasks, to elaborate a description of the system, and to identify information and action flows within the system [4]. Task analysis has many potential applications, including system design, system evaluation, training design and evaluation, interface design, job design, personnel selection, and system reliability analysis.

Cognitive Task Analysis (CTA) is a relatively recent outgrowth of general task analysis methods. *CTA* refers to a group of techniques used to capture and represent the cognitive elements underlying performance of a given task. CTA recognizes that, increasingly, automation in complex systems is changing the nature of work, and shifting emphasis from physical tasks (such as pushing buttons or pulling levers) to more cognitive tasks (e.g. monitoring, interpreting, analyzing, planning, diagnosing, deciding, etc). As a result of this shift, much of current-day "work," from air traffic control rooms, to nuclear power plants, is not directly observable. CTA was therefore developed to extend task analysis methods to the mental skills and processes (e.g. critical decisions) underlying observable behavior. CTA has been applied in various domains, from flight deck operations [5], to ATC [6], to military command and control [7], nuclear power plant operation [8] and process control [9]. Although there is some disagreement in the field, CTA is often used to decompose **both** the cognitive and behavioral aspects of task performance. CTA is particularly useful when the task involves elements of the following [10;11]:

- Complexity;
- Uncertainty;
- Decision making;
- Dynamic interactions; and
- Teamwork

2 Method

A functional task description was developed in a previous Functional Hazard Analysis (FHA), and this served as the basis for the CTA. Figure 1 shows a functional description of the overall analysis, into which CTA fit. On the basis of FHA, CTA aimed to identify the cognitive elements underlying performance in the RA scenarios, and to identify potential error mechanisms. CTA would in turn specify potential human errors, and thereby drive a Human Reliability Assessment (HRA) that sought to identify the probabilities of specific errors, by identifying such factors as controller reaction times, types of detection failures, interpretation errors, and potential controller workload issues.

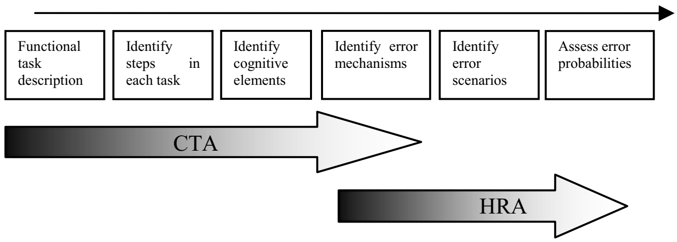


Fig. 1. Analysis of the RAD operational concept

Based on the FHA output, five interesting RA scenarios were selected. The five scenarios were all based on typical two-aircraft en-route encounters, in which an RA is presented to both air and ground, and disregarded such contextual factors as mixed equipage (i.e. where an intruder is not TCAS equipped), multiple aircraft encounters, etc. As shown in table 1, each of the five scenarios was considered in the context of two operational concepts: current day and RAD variants.

Table 1. The ten scenario variants

Scenario	**Operational Concept**	
	Current day	*RAD*
1: Nominal RA: The pilot reports correctly and in a timely fashion, and follows the RA correctly	**1**	**1R**
2: No report: No pilot report, but he correctly follows RA	**2**	**2R**
3: Incorrect report: Incorrect pilot report, but he correctly follows RA	**3**	**3R**
4: No report, no maneuver: No pilot report, and no RA	**4**	**4R**
5. Correct report, no maneuver: Pilot reports correctly, but does not maneuver	**5**	**5R**

Data collection for the CTA was conducted during one half-day session (and follow-up teleconference) between one researcher and a licensed air traffic controller. The controller was sent a packet of introductory material ahead of time on the RAD operational concept. Data collection began with a follow-up briefing on this material. The first step in the CTA was a card sorting exercise, in which potential tasks and sub-tasks were laid out in logical order. Second, a standard Hierarchical Task Analysis (HTA) was conducted to decompose the tasks as much as possible. Third, each of the five scenarios (x 2 variants each—each with and without RAD) were stepped through in logical order. During this exercise, the controller was encouraged

to think aloud about what information was required, from where the information came, the mental and physical steps involved, potential sources of error, etc. A series of prompt questions, and a list of potential cognitive elements, was used to guide discussion. An audio recording was made of the CTA session, and was later transcribed for analysis.

On the basis of the functional task description, tasks were then decomposed by typical hierarchical task analysis (HTA), in which complex tasks are systematically broken down into constituent elements (sub tasks). Tasks were decomposed into 3-6 subtasks, as recommended in [13]. A mental walkthrough was conducted to identify (based on prompt questions, as needed) the following types of information:

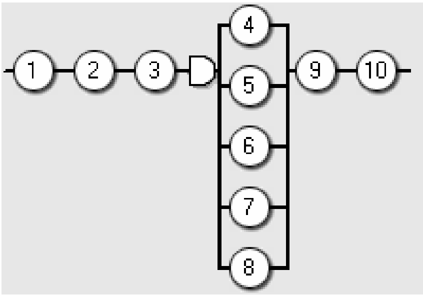
- Who was responsible;
- What, if any, decision was required at the moment;
- What information was required;
- Whether long term knowledge was required;
- What information was exchanged;
- Sequence dependencies in the information flow (e.g. “if he doesn’t give me that information now, I need this information then…”)
- Situation assessments (“here’s what is happening…”)
- Plans for possible contingencies (“if he doesn’t do this, I’ll do that…”); and
- Potential errors at each step.

On the basis of notes, sketches, and transcribed discussion, a formal CTA flowchart was developed for the two operational concepts (i.e. with and without RAD) under nominal conditions. Except where decomposition revealed relevant differences between the operational concepts, task description was kept as high as possible for the sake of clarity. The impact of both non-nominal events, and other “contextual factors,” was then examined with respect to their differential impact on the two operational concepts.

3 Results

The CTA focused on the period between when an RA is activated, through when the controller detects the RA (generally via pilot report or RAD, but sometimes on the basis of noticing a maneuver) until when ATC ultimately resumes authority. As shown figure 2, a total of ten high level tasks fell out of the CTA. In the process of normal scanning and control, the steps of confirming an RA encounter, confirming a deviation and ceding authority to the aircraft are carried out sequentially. Once authority is transferred to the aircraft, the controller then proceeds to carry out the remaining five tasks in parallel—monitoring for third party conflicts, providing traffic advisories, etc, all the while watching for signs (either via report or a return to clearance) that the RA encounter is complete.

Detailed presentation of results (expanded flowcharts ran into many pages) is well beyond the scope of this article. Instead, summary findings are presented here. On the



1 Confirm RA encounter
2 Confirm deviation
3 Cede authority to aircraft
4 Monitor RA encounter
5 Identify third party conflicts
6 Prevent third party conflicts
7 Provide traffic info to aircraft
8 Provide traffic info to other acft
9 Determine end of RA encounter
10 Resume ATC authority

Fig. 2. CTA-identified high level tasks, for the nominal RA scenarios

basis of the CTA, a number of potential issues came to light, with respect to the RAD concept. Among these are the following notable examples:

- That the chief benefit of RAD appears to lie in the speed with which changes (i.e. aircraft evasive maneuvers) can be anticipated and traffic located onscreen;
- That RAD can introduce potentially ambiguous control situations between air and ground;
- The RAD visual " pop-out" effect transforms some current ATC tasks (scanning screen, locating aircraft, etc) to largely perceptual ones;
- RAD could prime controllers to mistakenly hear what they expect to hear;
- Given the near certainty of RAD false alarms, the issue of trust must be addressed further.

4 Discussion

One chief conclusion from the CTA was that RAD can benefit both the speed and accuracy of locating aircraft onscreen, by transforming the current-day task of locating aircraft (remembering call sign, scanning screen, identifying aircraft calling) to a largely perceptual one. Specifically, it can do this by [1] preparing ATC to expect a report, and [2] by helping ATC determine (from the visual "pop-out" nature of RAD) where onscreen RAD aircraft are located.. In the absence of RAD, ATC must hear a call sign, locate onscreen the aircraft calling, and identify the intruder. RAD thus potentially benefits the early stages of an RA encounter, when ATC must confirm the presence of both an RA and the presence of a coordinated maneuver-- but it also pays potential dividends during and at the end of the RA encounter. However a few caveats are in order:

- RAD might prime ATC to hear what they expect to hear, and as a result mishear the subsequent pilot report;
- Despite the fact that a pilot report is necessary for ATC to cede authority, it seems that ATC will provisionally transfer authority on the basis of an RAD, and seek to gather confirmatory evidence of an RA;

- In the absence of such evidence, an ambiguous control situation can emerge either at the beginning or end of an RA encounter. In fact, pilots are now trained to report “Unable TCAS" in response, so the risk lies not in the aircraft following a clearance after RA, but in ATC mistakenly thinking it has authority, and planning other traffic accordingly;
- The timing of RAD can prompt ATC to query at the same time as pilot reports are to be expected, and can increase the chance of a “stepped-on´ transmission from encounter aircraft;
- There will be false/nuisance RAs. It is difficult to analytically determine the influence of trust in ATC’s willingness to (perhaps mistakenly) believe RAD in the absence of other evidence (e.g. maneuver, report);
- The potential costs of RAD in terms of attention-tunneling, and the risk of neglecting other traffic, must be clarified.

CTA was a judged afterward to have been an extremely valuable exercise. This was not a foregone conclusion, given the (deceptively) simple scenario apparent in the TCAS RA scenario. In retrospect, it is only by systematically stepping through such scenarios and identifying underlying information flows and decision points, that certain tasks, decision voids and error possibilities can become apparent.

References

[1] Kirwan, B., Ainsworth, L.K. (eds.): A Guide to Task Analysis. Taylor and Francis, London (1992)
[2] Beevis, D., et al.: Analysis techniques for man-machine system design. In: NATO document AC/243 (Panel 8)TR/7. vol. 1(2), Brussels, NATO, Belgium (1994)
[3] Schraagen, J.M.C., van Chipman, S.F., Shalin, V.L.: Cognitive Task Analysis. Lawrence Erlbaum Associates, Inc. Mahwah, NJ (2003)
[4] Stammers, R.B., Carey, M.S., Astley, J.A.: Task analysis. In: Wilson, J.R., Corlett, E.N. (eds.) Evaluation of Human Work, Taylor and Francis, London (1990)
[5] Latorella, K.A., Pliske, R., Hutton, R., Chrenka, J.: Cognitive task analysis of business jet pilots’ weather flying behaviors: Preliminary results. In: NASA-TM-2001-211034, NASA Langley Research Center, Hampton, VA (2001)
[6] Manning, Carol, A.: Measuring Air Traffic Controller Performance in a High-Fidelity Simulation. In: FAA technical report, FAA CAMI, Oklahoma City, USA (January 2000)
[7] Petersen, B., Stine, J.L., Darken, R.P.: Eliciting knowledge from military ground navigators. In: Montgomery, H., et al. (ed.) How professionals make decisions: Research and applications, pp. 351–364. Erlbaum, Mahwah, NJ, USA (2005)
[8] Rasmussen, J.: Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering. North-Holland, New York (1986)
[9] Vicente, K.J.: Cognitive work analysis: Toward safe, productive, and healthy computer-based work. Lawrence Erlbaum & Associates, Mahwah, NJ (1999)
[10] Gordon, S.E., Gill, R.T.: Question probes: A structured method for eliciting declarative knowledge. AI Applications In. Resource Management 3, 13–20 (1997)
[11] Reynolds, R.T., Neville, K.: Job Analysis versus Cognitive Task Analysis of Air Traffic Controllers. In: Paper presented at the Military Psychology Conference, VA (2002)

Glossary

ATC	Air Traffic Control
CTA	Cognitive Task Analysis
FARADS	Feasibility of ACAS RA Downlink Scenario
FHA	Functional Hazard Analysis
HRA	Human Reliability Analysis
RA	Resolution Advisory
RAD	Resolution Advisory Downlink
TCAS	Traffic Collision Avoidance System

The Development of a Cognitive Work Analysis Tool

Daniel P. Jenkins[1], Neville A. Stanton[1], Paul M. Salmon[1], Guy H. Walker[1], Mark S. Young[1], Ian Whitworth[2], Andy Farmilo[2], and Geoffrey Hone[2]

[1] School of Engineering and Design, Brunel University, Uxbridge Middlesex, UB8 3PH
[2] Department of Information Systems, Cranfield University at the UK Defence Academy, Shrivenham, SN6 8LA, UK
daniel.jenkins@brunel.ac.uk

Abstract. Due to their complexity, systems used within Network Centric Warfare and Command and Control are notoriously difficult to predict. These systems are often influenced by an ever increasing number of dynamic constraints. This dynamic instability causes problems for many traditional normative Human Factors techniques. Cognitive Work Analysis (CWA) is a formative process that focuses on these constraints rather than prescriptive methods of working; this constraint based approach allows the model to handle the unexpected and unanticipated events common in network-centric warfare. This paper presents the development of a Cognitive Work Analysis (CWA) software tool. The tool has two main purposes. The primary purpose is to assist the user in developing the large number of graphical representations that support the iterative design process. The secondary purpose is to explain CWA to novices and systematically guide them through the analysis process. The paper provides a brief introduction to CWA along with a description of the tool and its current capabilities.

Keywords: Cognitive Work Analysis; Tools; Software Development; Training.

1 Introduction

Cognitive Work Analysis (CWA) is a structured framework for considering the development and analysis of complex socio-technical systems. The framework leads the analyst to consider the environment the task takes place within and the effect of the imposed constraints on the system's ability to perform its purpose. The analysis guides the analyst through the process of answering the questions: *why* the system exists, *what* activities are conducted within the domain as well as *how* this activity is achieved, and *who* is performing it. Fidel & Peijtersen (2005) describe CWA as focusing simultaneously on: the task actors perform, the environment in which it is carried out, and the perceptual, cognitive, and ergonomic attributes of the people who conduct the task.

This constraint based approach separates CWA from many other Human Factors methods currently in wide use; according to Sanderson (2003) "CWA does not focus on how human-system interaction should proceed (normative modelling) or how human-system interaction currently works (descriptive modelling). Instead, it focuses on identifying properties of the work environment and of the workers themselves that

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 504–511, 2007.

define possible boundaries on the ways that human-system interaction might reasonably proceed without explicitly identifying specific sequences of actions (formative modelling)." Naikar (in press) points out that by focusing on constraints, rather than on particular ways of working, CWA aims to support workers in adapting their behaviour online, in real time within system constraints, allowing them to maintain performance and safety in a variety of situations including unanticipated events. Naikar and Lintern (2002) describe the framework as supporting revolutionary rather than evolutionary design; rather than focusing on how we use new technologies to execute our current tactics and doctrine better, CWA, through its focus on constraints, aims to determine how the new technologies enable us to do things differently. Traditional methods allow evolutionary incremental improvement. It is only through formative thinking based on approaches such as CWA that exponential improvements can be realised.

The CWA framework is attracting increasing levels of interest in a wide range of socio-technical domains including, aviation (e.g., Naikar & Sanderson, 2001); process control (e.g., Vicente, 1999); nuclear power (e.g., Olsson & Lee, 1994); Naval (e.g., Bisantz et al, 2003); military command and control (e.g., Jenkins et al, in press); road transport (e.g., Salmon et al, In Press); health care (e.g., Miller, 2004); air traffic control (e.g., Ahlstrom, 2005); and manufacturing (e.g., Higgins, 1998). The framework has been used in system modelling (e.g. Hajdukiewicz, 1998); system design (e.g. Bisantz et al, 2003); training needs analysis (e.g. Naikar & Sanderson, 1999), training program evaluation and design (e.g. Naikar & Sanderson, 1999); interface design and evaluation (Vicente, 1999); information requirements specification (e.g. Ahlstrom, 2005); tender evaluation (Naikar & Sanderson, 2001); team design (Naikar et al, 2003); and error management strategy design (Naikar & Saunders, 2003).

By using a suite of tools (see Fig.1) the framework models the system from five different perspectives defined by Vicente (1999).

According to Naikar (2006) the first phase, Work Domain Analysis, identifies the constraints on workers' behaviour that are imposed by the purposive and physical context, or problem space, in which workers operate. Work Domain Analysis is used to define the task environment. It identifies a fundamental set of constraints on the actions of any actor, thus providing a solid foundation for subsequent phases. The system domain is represented at a number of conceptual levels. At the highest level the system's raison d'être is represented; at the lowest level the physical objects within the system are described.

The second phase, Control Task Analysis, is used to understand activity within the domain; it allows us to identify the requirements associated with known, recurring classes of situations. This phase specifies the input and end goal leaving a 'black box' in the middle. The phase identifies what needs to be done independently of how or by whom.

Strategies Analysis, the third phase of CWA looks at filling in the 'black box' left in control task analysis; it looks at different ways of carrying out the same task. Wherever the previous phase dealt with the question of *what* needs to be done this phase addresses *how* it can be done.

The fourth phase, Social Organisation & Cooperation Analysis, addresses dividing the task between the available resources and looks at how the team communicates and

cooperates. The objective is to determine how the social and technical factors in a socio-technical system can work together in a way that enhances the performance of the system as a whole.

The fifth and final phase, Worker Competencies Analysis, focuses on the types of behaviour required by workers to conduct a task. This section addresses the more traditional core concerns of the Human Factors and HCI communities.

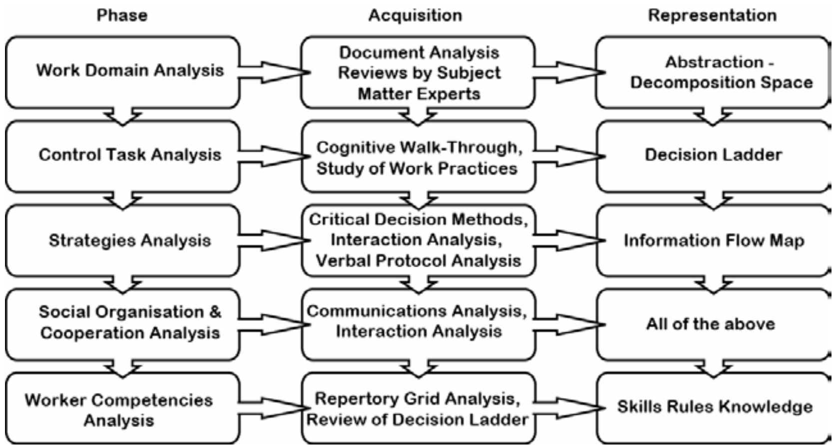


Fig. 1. The five phases of CWA according to Vicente (1999) (Acquisition methods added from Lintern et al (2004))

2 The Tool

The CWA process is often criticised for being complex and time consuming (Cummings, 2006); the tool presented in this paper attempts to provide some level of guidance and expedite the documentation process in each of the five phases identified by Vicente (1999) to address these concerns. In order to provide guidance the tool uses an example scenario of an experimental sensor to effecter paradigm (Jenkins et al, in press) to describe and illustrate the process.

The approach taken by an analyst conducting a CWA is highly dependent on the domain and the aims of the research. Many CWA analyses will not focus on all of the five phases; the majority of analyses place a much greater focus on the first phase, Work Domain Analysis. The type of analysis conducted is likely to depict which of the phases are used and in what order and ratio. Interface design such as ‘Ecological Interface Design’ (EID) (Burns & Hajdukiewicz, 2004) tends to use only the first phase (Work Domain Analysis) when designing new interfaces. An analysis of team design or training needs is more likely to focus more heavily on Control Task Analysis and Social Organisational and Cooperation Analysis as these phases capture activity and consider its distribution amongst the system’s assets. The benefit of the framework is that the CWA approach can be applied throughout the system life cycle for the design, development, representation and evaluation of both conceptual and existing operational systems. In order to support this often non-linear process the tool

has been designed with a tabbed bar built into the interface, this tabbed bar allows novices to approach the framework in a linear process whilst allowing users to jump backwards and forwards at their own convenience. Each of the phases is decomposed into a number of sections (see Fig. 2). The first section introduces the phase in generic terms briefly explaining the principles behind the particular level of the framework. The next section provides a worked example of the analysis at the current phase consistently using the sensor to effecter case study (Jenkins et al, in press) The example is fully annotated in an attempt to provide guidance to novices, and to provide an idea of the finished product. The subsequent sections allow the user to construct and develop the models and documentation required; the process is supported by mouse-overs and prompts to guide novices in creating diagrams.

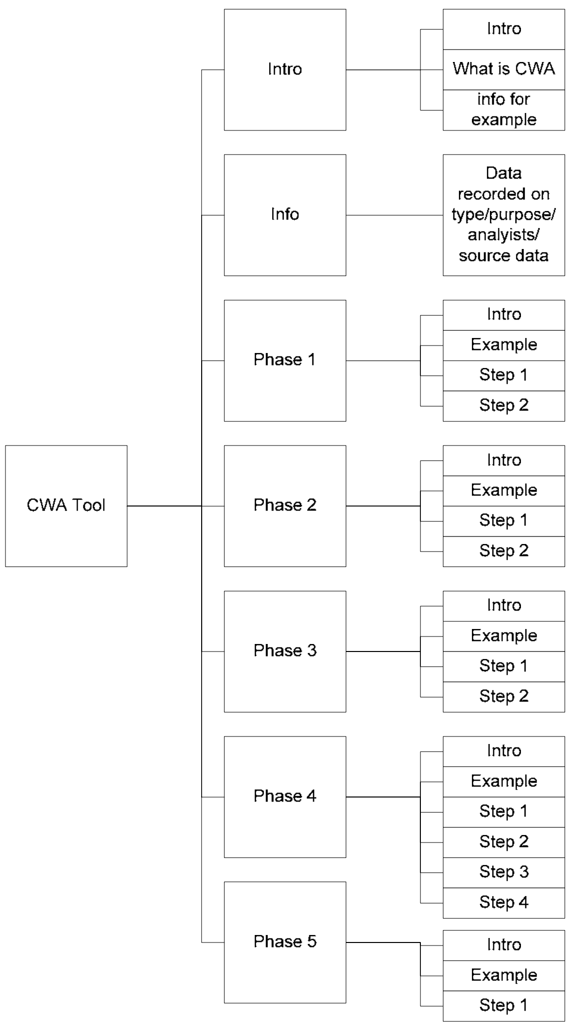


Fig. 2. Map of tool function

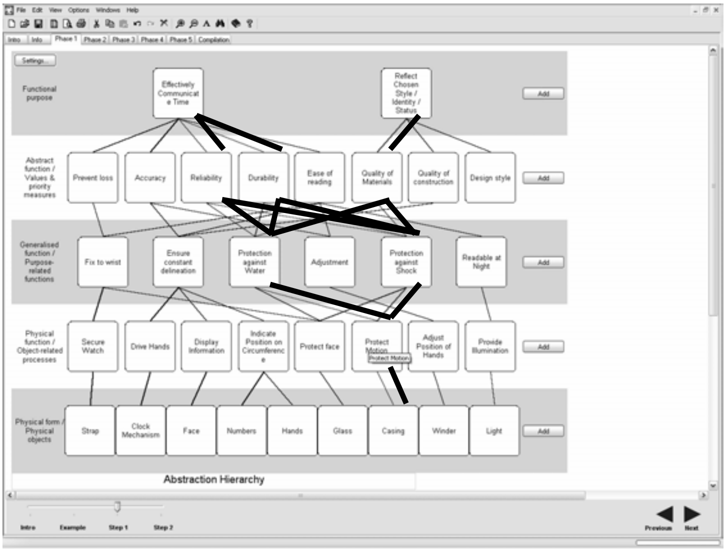


Fig. 3. Example of Abstraction Hierarchy created by the tool

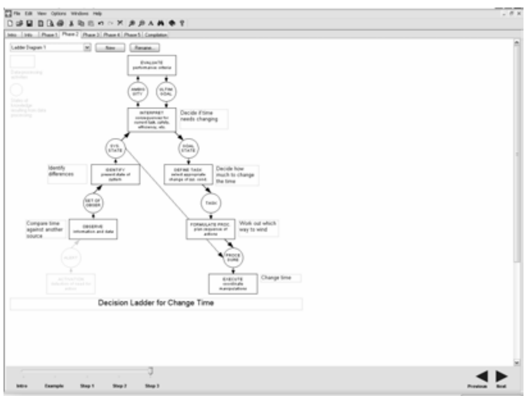


Fig. 4. Example of Contextual Activity Template created by the tool

The dynamic nature of the software tool has many benefits over its paper based counterpart; allowing files to be saved, copied, edited and rapidly transmitted. Feature such as an ability to export all of the products to a Word document with a single button click also make the report compiling process much simpler. Diagrams are made far more manageable by the tool facilitating the ability to rapidly reorder the diagrams and representations. The tool has been intentionally designed to be unconstrained allowing researches to apply their own interpretations of the framework, although in order to constrain novices from making fundamental mistakes some of the functions are deliberately not permitted, such as connecting to data processing activities in the decision ladder or connecting a node in the abstraction hierarchy to a node two levels above.

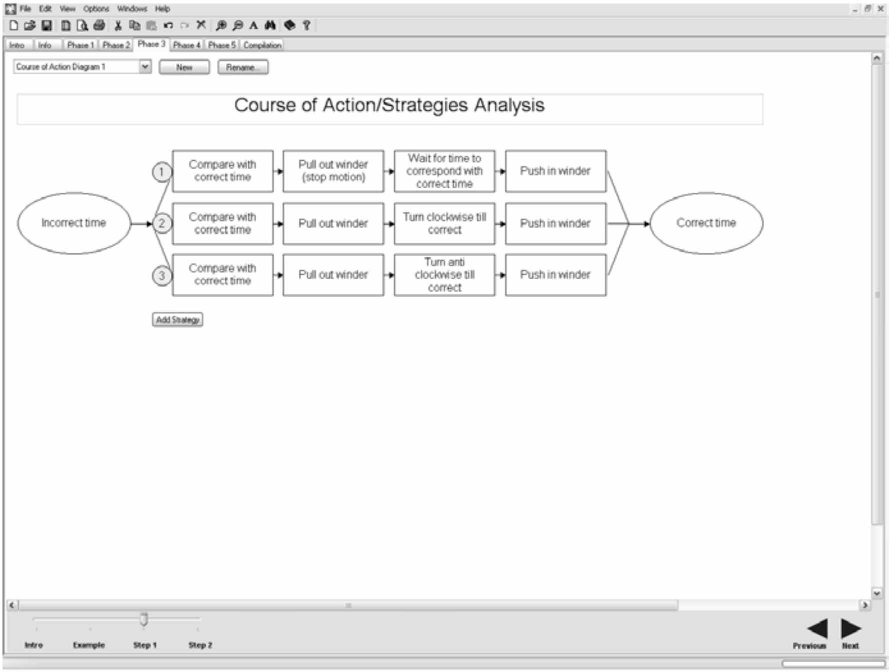


Fig. 5. Example of Strategies Analysis created by the tool

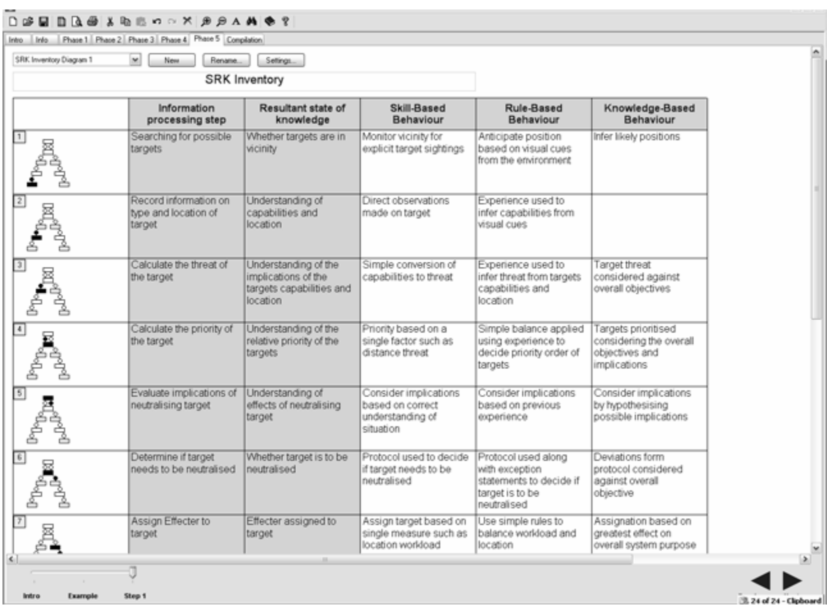


Fig. 6. Example of Skills Rules and Knowledge Table created by the tool

Additional benefits from the dynamic nature of the tool have also been exploited: When the user places their mouse over a node on the Abstraction Hierarchy the linked nodes both up and down the hierarchy are highlighted in red (for illustrative purpose these have been shown in bold in); this allows a cause and effect relationship to be examined within the different levels. The speed in which the abstraction hierarchy can be developed, edited and reviewed (helped by the addition of the red links) makes it feasible to carry out the generation and validation of representation from scratch with Subject Matter Experts (SMEs) in real time.

The software tool passes important data forward aiding the completion of subsequent representations. This significantly reduces the tedium of the documentation process. This also means that minor changes to text semantics and diagram layout is fed through from the initial stages to the subsequent phases. This has particular benefits in the Social Organisation and Cooperation phase, which reuses the products generated in the previous three phases. This semi structured process combined with the ease of documentation further encourages analysts to continue their analysis beyond the first two phases of CWA, a concern raised by Cummings (2006).

3 Conclusions

This paper has briefly introduced CWA along with a short description of the software tool and its potential benefits.

Whilst the tool is still in its development stages some of the benefits in terms of novice training and its ability to expedite the documentation process are already clearly visible. The dynamic nature of the tool has also revealed some unanticipated benefits in eliciting domain information from subject matter experts.

The development of the tool now continues with a global structured feedback process led by some of the leading researchers in CWA.

Acknowledgement. This research from the Human Factors Integration Defence Technology Centre was part-funded by the Human Sciences Domain of the UK Ministry of Defence Scientific Research Program.

Any views expressed are those of the authors and do not necessarily represent those of MOD or any other UK government department.

References

1. Ahlstrom, U.: Work domain analysis for air traffic controller weather displays. Journal of Safety Research 36, 159–169 (2005)
2. Bisantz, A.M., Roth, E., Brickman, B., Gosbee, L.L., Hettinger, L., McKinney, J.: Integrating cognitive analyses in a large-scale system design process. International Journal of Human-Computer Studies 58, 177–206 (2003)
3. Burns, C.M., Hajdukiewicz, J.R.: Ecological Interface Design. CRC Press, Boca Raton, FL (2004)
4. Cummings, M.L.: Can CWA inform the design of networked intelligent systems. Moving Autonomy Forward Conference 2006, Lincoln, UK (2006)

5. Fidel, R., Pejtersen, A.M.: Cognitive Work Analysis. In: Fisher, K.E., Erdelez, S., McKechnie, E.F. (eds.) Theories of information behavior: A researcher's guide, Information Today, Medford, NJ (2005)
6. Hajdukiewicz, J.R.: Development of a structured approach for patient monitoring in the operating room. Masters Thesis. University of Toronto (1998)
7. Higgins, P.G.: Extending cognitive work analysis to manufacturing scheduling. In: Proceedings of OzCHI'98, November 30-December vol. 4, pp. 236–243 (1998)
8. Jenkins, D.P., Stanton, N.A., Salmon, P.M, Walker, G.H., Young, M.S.: Cognitive Work Analysis of a Sensor to Effecter System: Implications for Network Structures. Human Factors and Ergonomics Society Europe Chapter Meeting 2006 (in Press)
9. Lintern, G., Cone, S., Schenaker, M., Ehlert, J., Hughes, T.: Asymmetric Adversary Analysis for Intelligent Preparation of the Battlespace (A3-IPB) United States Air Force Research Department Report (2004)
10. Miller, A.: A work domain analysis framework for modelling intensive care unit patients. Cognition, Technology and Work. 6(4), 207–222 (2004)
11. Naikar, N.: Modelling activity in network-centric operations with cognitive work analysis: Work situations, work functions, decisions, and strategies. In: Bolia, B. (ed.) Supporting decision effectiveness in network-centric operations. Wright Patterson Air Force Base, Dayton, OH (in press)
12. Naikar, N.: Beyond interface design: Further applications of cognitive work analysis. International Journal of Industrial Ergonomics 36, 423–438 (2006)
13. Naikar, N., Lintern, G.: A review of Cognitive Work Analysis: Towards safe, productive, and healthy computer-based work by Vicente, K.J. The International Journal of Aviation Psychology, 12(4), 391–400 (2002)
14. Naikar, N., Sanderson, P.M.: Work domain analysis for training-system definition. International Journal of Aviation Psychology 9, 271–290 (1999)
15. Naikar, N., Sanderson, P.M.: Evaluating design proposals for complex systems with work domain analysis. Human Factors 43, 529–542 (2001)
16. Naikar, N., Saunders, A.: Crossing the boundaries of safe operation: A technical training approach to error management. Cognition Technology and Work. 5, 171–180 (2003)
17. Olsson, G., Lee, P.L.: Effective interfaces for process operators. The. Journal of Process Control 4, 99–107 (1994)
18. Salmon, P.M., Stanton, N.A., Regan, M., Lenne, M., Young, K.: Work domain analysis and road transport: Implications for vehicle design. International Journal of Vehicle Design (In Press)
19. Sanderson, P.M.: Cognitive Work Analysis across the system life-cycle: Achievements, challenges, and prospects in aviation. In: Pfister, P., Edkins, G. (eds.) Aviation Resource Management (Vol 3). Aldershot, UK Ashgate (2003)
20. Vicente, K.J.: Cognitive work analysis: Toward safe, productive, and healthy computer-based work. Lawrence Erlbaum Associates, Mahwah, NJ (1999)

Human Activity Modeling for Systems Design: A Trans-Disciplinary and Empirical Approach

Saadi Lahlou

EDF R&D/SINETICS/ Laboratory of Design for Cognition
1 ave du Gl. De Gaulle,
92141 Clamart, France
CNRS-EHESS/IIAC/ Centre d'Etudes Transdisciplinaire,
UMR 8177, Paris, France
saadi.lahlou@edf.fr

Abstract. Complex system design is about technical specifications, but also how humans act in the loop in Dissemination, Operation, Maintenance and Evolution (DOME) of the system. In system specification, we focus on distribution of work between humans and systems. In design of system's DOME, on building an ecology of individual motives, organizational rules and mediating structures to keep the system sustainable. In our participative design process itself, on how to enroll and maintain test users in realistic experiments. We adopt a complementarist approach: we use different models of human cognition and behavior for each aspect. Human behavior is determined by many factors including subject's motives and goals, constraints and affordances of the context. We list here which models we use and how.

Keywords: design**,** activity theory, cognitive attractors, cognition, social representation, affordances, goals, distributed cognition, situated action, ecological psychology, involvement.

1 Introduction

A main issue in complex systems design is that they include humans in their loops. Most devices and sub-systems are man-made artifacts; they are therefore somewhat predictable, even in their failure. But the "human factor" remains a source of uncertainty in design and operation. Of course, humans are also partly cultural artifacts, and behavior can be made predictable by training and rules. But there are limits. As Lomov [1963, p. 23] states: "(…) a man remains a man even when he enters the role of a link in the control system". We need humans in the loop of complex systems since Man is more plastic than machines with regard to information input, processing, execution, and is able to "grasp" improbable events [Lomov, 1963, p. 21]. So at least in the near future of complex systems design the human factors issue is here to stay.

After a brief overview of literature (section 2), we describe our design approach, "experimental reality" (section 3). *Experimental reality* is based on collaborative design and incremental change; it has lighter requirements in terms of cognitive

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 512–521, 2007.

models of the user. Section 4 describes which models of human behavior we use, for system functions specifications, for "process of system design" itself, for design of the system's life cycle: Dissemination, Operation, Maintenance and Evolution (DOME).

2 Models of Human Cognition

Many are the models of human cognition and behavior. Let us briefly mention some, in a short and incomplete overview, to show the variety of existing models, and how they address different levels: mind, subject, organization, culture.

In Herb Simon's seminal work [Simon, 1945, Simon & Newell, 1972], the human is seen as an entity who solves problems. Subjects strive to find an efficient pathway in an incompletely known "problem space". Using bounded rationality, they apply heuristics to reach solutions until they find one which is good enough ("satisficing"), e.g. by trying existing routines and previously known solutions.

In this now classic perspective, an important issue is what mental structures and which cognitive processes are used to represent the situation, and deal with it. *Mental models* [Johnson-Laird, 1993], *social representations* [Moscovici, 1961; Abric 1994], *simuli* [Minsky, 1985] *scripts* [Shank & Abelson, 1977, Rumelhart & Norman, 1983] are some the many theories dealing with these entities.

Other authors focus on how the situation provides actors with constraints and possibilities. *Ecological Psychology* describes how elements in the context are directly interpreted in terms of what action is possible: "connotations of activity" [Uexküll, 1965] or "affordances" [Gibson, 1967]. *Situated action* [Suchman, 1987; Lave, 1988] describes how action emerges as subjects try to realize their plans while been taken in the flow of situations. Subjects can make sense of their *course of action* [Theureau, 1992] but they are far from being in full control. Still, there are general principles which account for the activity of operators in complex systems.

Rasmussen (1983, 1985) with the *Skills, Rules, Knowledge* model introduces a hierarchical perspective on the levels of operator behavior, as problems require more or less cognitive resources to be solved (from the simple routines at "Skill" level to complex reasoning at "Knowledge" level for completely new situations).

Subjects are not solely driven by the problem: they have intentions, goals and values. *Activity theory* [Rubinshtein, 1940, Leontiev, 1959, 1974; Nosulenko & Samoylenko, 2006] provides powerful conceptual tools with the notions of goals (conscious representation of the desired state) and motives, and of the Man's relations with instruments [Rabardel, 1995]. In the course of goal-driven activity, "actions" are steps to the goal in the conditions given. Actions or parts of actions can become automatic "operations" which are executed beyond consciousness. "Subjects" in activity theory can be individual or collective.

The collaborative aspect of work is dealt with by many models grounded in anthropology, sociology and social psychology. These models tend to include institutions and artifacts as major factors for coordination. In these theories, the focus is on the system as a whole. Humans are considered from his perspective as agents in the system, while in the previous approaches the system was rather considered as a context for human action. *Cognitive sociology* (Cicourel, 1973) and *distributed*

cognition (Hutchins, 1995a; Cicourel, 1994) are powerful conceptual frameworks; they describe how processes are distributed over the cultural and socio-technical system, and how artifacts and institutions play the role of *mediating structures* (Hutchins, 1995) which guide action.

This short trip in theory shows how varied current models are; we believe they all do have some explicative value. This is why we opt for a complementarist approach [Devereux, 1970]. Instead of choosing a single model of cognition for everything[1], we use different models for different situations or design issues, (see section 4).

Although global theories may be quite different, a recurrent, mainstream, trend in cognitive ergonomics when it comes to application and actual systems design is to consider the human as an information processor and decision maker, with specific focus on *cognitive bottle-necks*, especially attention and memory. In this perspective, great care is taken to design interfaces for low workload, to avoid multitasking, to distribute information over several sensory modalities to avoid overload. The empirical literature reflects this pragmatic approach which goes with the need to give quantitative specifications to the engineers who build the devices and interfaces in the system. Operational models and methods exist which link theories of cognition to this technical issue of resource allocation and Man-System interface design, e.g. Cognitive Function Allocation [Boy, 2001].

In the design phase, a risk is that a great deal of the cognitive aspects of the aforementioned theories is lost, and that in the final system humans find themselves playing the role of the general-purpose plastic link between rigid mechanical systems described by Lomov. In other words, humans in complex systems do have a specific functional role, but they often also implicitly have the vicariant role of fixing a technical system that was built imperfect. This reality of this "human as a repair" contrasts with the claims that systems are user-centered. Our design approach tries to avoid this problem by constantly keeping users empowered in the design process, hereby forcing designers, engineers and other stakeholders to *continuously* take the human user's perspective into account.

3 Design in Experimental Reality

Constructing models of human cognition in general is a daunting task. Fortunately, designing systems is a specific activity and therefore we do not need to solve the general issue of human cognition.

Also, some design approaches need less cognitive modeling than others. This is the case of our design approach, e*xperimental reality,* based on procedural, incremental changes rather than on a full analysis and specification of the problem space. This section describes our approach.

"Experimental reality" [Lahlou, Nosulenko & Samoylenko, 2002] is an attempt to practice *participative design* in comfortable conditions. The general idea is to obtain the enrollment of a population of subjects in a long term realistic test of successive versions of the system being developed, under intensive observation. To do this, *a whole real system is instrumented for observation*. For example, when designing systems to fight against cognitive overload, we transformed individual offices into

[1] "For who has only a hammer, everything looks like a nail".

observation rooms with continuous monitoring, where we could change various elements of furniture and measure changes [Lahlou, 1999]; to design augmented work environment, we constructed a whole building which was equipped for observation, and then inside this living laboratory we installed a population of real users with their workstations to do their everyday job [Lahlou et al., 2002]. The experimental but real system becomes a “work in progress” where designers and users work *together* on a continuous basis to imagine and test new arrangements. For example, to develop augmented meeting rooms, we provided a comfortable meeting space which could be reserved for free, in an industrial facility housing more than 2000 office workers (engineers, scientists, administrative personnel). Maintenance and operation of the space was provided by the designer team, so the space was fully functional and used as such; hundreds of actual meetings took place in the space. This space was continuously filmed and interactions analyzed [Lahlou, 2005]. The design team could then modify gradually the system (here, the room, devices, interfaces etc.) to fix emerging issues. In doing so, the design team focuses attention on the failures of the system, on the parts of the process which the users complain about, or which seem “strange” or costly from an engineering or human factors perspective. This is a first place where we need cognitive models of activity. This approach is different from testing a new solution on a case, since our process also interweaves both test and continuous (re)design phases.

This design strategy needs strong and long term support from the management ; and trust from users ; it requires straight ethics so that observation is accepted by observees, it costs important investments in observation (sensors, procedures and analysis) ; but in the end it proves more efficient and cost-effective than classic approaches. This strategy is especially adapted to the design of complex systems which must be scaled out and endure continuous evolution because of technological drift. For example, in the case of augmented collaborative rooms, even though meeting rooms have been scaled out in the company since 2002 based on the specifications of the test room, we currently (2007) keep research on this “mother room” active in order to test new devices and possible improvements. When these improvements prove positive, theses changes are scaled out to the existing rooms, therefore ensuring that the whole fleet of rooms is up to date.

An interesting aspect of this approach is that it solves naturally the problem of transition from present state of the art to the new system; since what it does is precisely experimenting this transition with real users.

This design strategy can be applied only when one can construct a system that is operational enough to be used by actual users in a real process; and if this population of users can be enrolled in “playing the game” of continuous design. This is another place where we need cognitive models of humans. Although this seems quite a tough requirement, in fact these conditions are not as drastic as it may seem. Indeed, many complex systems are actually not designed from scratch, but often stem from a present version which can then be used as the test system. And recruiting users in the design process is also often possible as long as the design process accepts to take into account users’ goals and personal interests beyond the mere functional optimization of the system.

4 Models of Cognition for Experimental Reality

In order to proceed with experimental reality, we need models of human behavior at three levels: for system functions specifications, for design of DOME, for the process of system design itself.

In system specification phase we try to understand which functions can be supported by the devices, and which by humans; we also investigate the sources of malfunction, mistakes, and users' displeasure. We focus especially on how the environment can prompt and support the users into best collective practice.

In the design of DOME, we focus on the growth and sustainability of ecology of individual motives, organizational rules and more generally of mediating structures which will ensure the viability of the system on a daily basis at reasonable cost.

In the process of system design, we are interested in users as participants in the design process, we focus attention on how to enroll and maintain the relevant actors in the design process.

These three aspects are not independent. Some of the cognitive models we use serve more than one aspect, but they will be presented separately for more simplicity.

4.1 Design of Functional Specification of the System

The main model we use is Russian *activity theory*, because it helps specifying the goals of the system (common goals). We try to avoid situations where goals of users as individuals are against the common goals. So, at every step of design, we try to make *explicit* the goals of the humans in the loop. We do so by observing and asking the subjects. In the frequent cases where there is discrepancy between goals, we try to modify the situation so that individual goals can be pursued without conflicting with global system operation. It is our philosophy not to try to fight against individual goals because we consider that individuals will find their way through whatever counter-measures we could design as long as they are motivated; the final cost of their frustration or hidden activity to reach satisfaction may be more costly to the system than organizing legal or tolerated paths for individual satisfaction.

For example, many reasons why individuals participate in meetings have nothing to do with the meeting agenda. E.g. they want to discuss informally other issues with some participants [Lahlou, 2005]. Unless this goal can be satisfied in the margins of the meetings (e.g. onset, breaks) participants may be restless and tend to introduce theses topics in the discussion. This was a problem in the design of systems for distant collaboration (visio-conference, etc.) since the structure of distant-meetings breaks was not adequate for such informal discussions. We therefore installed a "boarding" procedure which enables informal contacts.

Understanding the goals and motives of operators is also important because these goals can be used to remind the operator what is important e.g. in emergency.

Another model we use is *cognitive attractors*. It states that humans are always potentially following several paths of activity, since they carry many goals to reach; and that subjects arbitrate their resources between these different paths. In other words, subjects continuously choose between various potential problem spaces to deal with ("problem choosing"). Empirical evidence suggests they tend to choose the problem with the best value/cost ratio. Cognitive attractors theory states "problem

choosing" is often a low level "capture" of the subject by a specific type of activity rather than the result of a conscious decision-making process.

Capture by a specific activity occurs when a critical mass of elements of this activity both in the context and in the mind "attracts" the subject. The force of the attractor is a combination of the value of the goal of the activity, the cognitive cost to be spent, and the pregnance of the elements [Lahlou, 2000].

When designing new systems, we manipulate these factors in order to keep the humans in the track of what they are supposed to do individually, to ensure a correct global operation of the system. For example, we use displays to keep operators focused on their present goal and task; we try to filter out potential distractors, we carve the affordances of devices so as to keep present course of action at a lower cognitive cost than potential distractors. Also we try to keep the activity *fluid* since break-downs or pauses in the process are often moments when subjects switch to another activity. For example a pause when the subject waits for system-response is often used to start another (short) task in parallel.

Ecological psychology is the cognitive model which guides us in the design of affordances of devices. We try to make "what is to be done" explicit in the form factor of objects. Moreover we do not merely consider that objects have affordances or connotations of activity; we consider them as full actors in the system, as does Latour [1993]. This means that not only devices should provide system feed-back to users, but also that they should behave as agents capable of initiative, e.g. by modifying their display to engage humans into specific actions. In this perspective we consider the system as an arena for set up for distributed activity (and this is how we understand Hutchins's distributed cognition), where an ecology of agents try to make the state of the system progress towards the goal [cf. Hutchins's cockpit, 1995b]. In this perspective, human or artificial agents reciprocally serve each other as mediating structures to guide, control, and remind what is to be done. Any part of the system which recognizes a specific situation or critical state through its sensors and interpretive equipment may signal it to other parts of the system.

Social representation theory [Moscovici 1961; Abric, 1994] helps us to understand how the common knowledge necessary for coordination between actors is structured, created, and disseminated. Social representations serve as an internalized user's manual for operating objects and behaving in the system [Lahlou, 2001]. Creating or modifying representations, and pushing the subjects into mobilizing the proper ones, is a crucial aspect of fluid system operation. For example, single participants in multiplex visio-conference meetings behave differently whether they participate from their office or from a meeting room. When they are in a meeting room, the social representation of "meeting" is activated by the context and they tend to be more attentive and to sidetrack less in individual tasks (like processing their email). A new practice goes with a modified social representation, and therefore constructing this modified representation (with training, manuals, rules, communication etc.) is part of interaction design.

Still, in this ecological vision of a collaborative work between humans and other artifacts, our bias is towards the anthropocentric perspective adopted by Russian psychology (Leontiev, 1974). Although artifacts and humans are involved in a collaborative activity, humans have preeminence in decision and the right to bypass or override mechanical systems.

4.2 Design of Dissemination, Operation, Maintenance and Evolution

A system must survive, and evolve in time to do so since its environment changes. Dissemination, Operation, Maintenance and Evolution take place in every system. If they are not designed *ab initio*, the system will not adapt properly to its environment, and the organization will have to input pressure and investments to keep it alive. For example, dissemination can cost a lot if the system is difficult to learn, and even more if it conflicts with other existing systems or practices. Such costs often stay hidden because they incur on other budgets, but from the global organization and the user's perspective, operation is much more fluid and comfortable is these issues have been considered and solved at initial design stage.

First, we check every stake-holder will get a benefit in having the system work. This is true for the innovation phase, as Latour shows, but also for the rest of the life of the system. For every problem encountered, we set up a solution which brings some benefits to the stake-holder involved. Some of these solutions may very pragmatic and involve managerial decisions and trade-offs (e.g. recognizing a new qualification to employees with a specific pay bonus); for the dissemination phase we may set up informal deals with specific individuals (e.g. contributing to the test brings them in contact with new partners or other Divisions, which may help their career).

In Dissemination design, when we use trickle down theory [Veblen, 1899]. We make participation in the new system desirable and fashionable by involving the top management. Another technique we set up with success it to build in some empowerment for viral dissemination. E.g. for our collaborative platforms, any registered user of the system can enroll another user upon his own responsibility and therefore get the social benefit of this offering this "enablement" Lahlou, 2005].

But finally, the most important is to keep users in track, by manipulating costs and creating social rules so as to make the "good practice" (understand: the one coherent with the common goal) have a lower cognitive cost for the user than other practices. This is done, as said in section 4.1 by distributing mediating structures over the environment to remind the users the goals, provide them with adapted affordances all along the activity path, filtering out potential "distracters", and organizing "drainage" so that individual goals can be satisfied in way that does not get in the way of general goals. So what we try to design is not a system-as-it-should-be, but a system-in-actual-use, including the "informal" aspects inherent to the fact that Man-in-the-loop is a human, with other interests and goals than simply be a part of *this* system. To start with, each human is part of *many* systems inside and outside the organization.

As no present cognitive theory in powerful enough to directly provide specifications for all these aspects, we proceed by trial and error during design phase and all along the system's life cycle, by observing and solving problems and discrepancies as they emerge, in a pragmatic way. Keeping a trustful and constructive relationship with the users provides us with rapid feedback and often with excellent design ideas. When users believe that reporting problems and suggesting solutions will indeed produce re-designs of the system, they get motivated in doing so and their contribution prove extremely useful. Not all users have good ideas all the time, but some users have very good ideas sometimes.

4.3 Models for the Process of System Design Itself

Our approach depends upon the involvement of the users in the design process. For this, we rely again on activity theory: we try to find ways of aligning the test user's goals with ours. For example, we ensure that users have some real interest (and no risk) in providing us with accurate behavioral data, and to use our experimental systems. This may lead us to choose subjects who either have a strong interest in the new functionalities offered by the new system; or have a specific interest in showing cooperation, or are interested in using the behavioral data for their own purpose. This is especially true in the first stages of design when the advantages of the new system are counterbalanced by poor usability, and when we need "friendly users".

Another theory we use is the social psychology of involvement [Lewin, 1952; Joule & Beauvois, 1998]. E.g. subjects get more involved when they have top "pay" in some way. We make the participation of a user a volunteer and formal process, where users sign off "informed consent". Subject's involvement is public, and ritualized. They are warned that having access to the experimental system is a privilege, in exchange of which they may have to undergo cumbersome interviews. Everything is done to transform the participation into a formal social contract between the subjects and the design team, especially in the initial phase.

5 Conclusion: Better Many Conceptual Tools Than One

Our pragmatic design approach led us to abandon the procrustean dilemma of choosing a one-fits-all model of cognition. Although our position is sometimes difficult to stand on academic grounds, we use different theories for different aspects of our design process, mainly: *activity theory, cognitive attractors, distributed cognition, ecological psychology, social representations, social psychology of involvement*.

Most important remains to set up a work-in-progress structure where actual users are empowered along the whole design and DOME process. This way, once again, the human factor brings innovation and compensates for the limitations of models. The fact that our design teams mix engineers, designers, stake-holders and end-users helps us to stay focused as a group on the problem, and to keep models of cognition in the modest role of *tools*. When we try to create solutions our belief in the power of these models remains limited. Experience showed that, in practice, it is sometimes possible to solve, *in specific cases,* problems for which we do not have, yet, general solution in theory.

References

1. Abric, J.-C.: Pratiques sociales et représentations. Paris: P.U.F (1994)
2. Boy, G.A.: Cognitive function analysis for human-centered automation of safety-critical systems (1998). In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Los Angeles, CA, pp. 265–272 (1998)
3. Cicourel, A.V.: La sociologie cognitive (1973). Paris: P.U.F, 1979 (1973)
4. Cicourel, A.V.: La connaissance distribuée dans le diagnostic médical. Sociologie du travail (4): 427–449 (1994)

5. Devereux, Georges.: Essais d'ethnopsychiatrie générale. Gallimard, Paris (1970)
6. Gibson, J.J.: Notes on affordances (1967). In: Reed, E., Jones, R. (eds.) Reasons for realism. Selected Essays of James J. Gibson. Lawrence Erlbaum Associates, London, pp. 401–418 (1982)
7. Hutchins, E.: Cognition in the Wild. MIT Press, Cambridge, MA (1995a)
8. Hutchins, E.L.: How a cockpit remembers its speed. Cognitive Science 19, 265–288 (1995b)
9. Johnson-Laird, P.N.: La théorie des modéles mentaux (1993). In: Ehrlich, M.F., Tardieu, H., Cavazza, M. (eds.) Les modéles mentaux. Approche cognitive des représentations, pp. 1–22. Masson, Paris (1993)
10. Joule, R.V., Beauvois, J.L.: La soumission librement consentie. P.U.F, Paris (1998)
11. Lahlou, S.: Observing Cognitive Work in Offices (1999). In: Streitz, N.A., Hartkopf, V. (eds.) CoBuild 1999. LNCS, vol. 1670, pp. 150–163. Springer, Heidelberg (1999)
12. Lahlou, S.: Attracteurs cognitifs et travail de bureau. Intellectica 2000/1 30, 75–113 (2000)
13. Lahlou, S.: Functional Aspects of Social Representations. In: Deaux, K., Philogene, G. (eds.) Representations of the Social, pp. 131–146. Blackwell, Oxford (2001)
14. Lahlou, S.: Cognitive Attractors and Activity-Based Design: Augmented Meeting Rooms. Human Computer Interaction Int'l 2005, Las Vegas, NA. vol. 1 (2005)
15. Lahlou, S., Nosulenko, V., Samoylenko, E.: Un cadre méthodologique pour le design des environnements augmentés. Social Science Information 41(4), 471–530 (2002)
16. Latour, B.: Nous n'avons jamais été modernes (1999). In: Essai d'anthropologie symétrique, La Découverte, Paris (1991)
17. Lave, J.: Cognition in Practice. Cambridge U. P, Cambridge (UK) (1988)
18. Leont'Ev, A.N.: The problem of activity in psychology (1974). Soviet Psychology 13(2), 4–33 (1974)
19. Leontiev, A.: Le développement du psychisme (3éme éd., 1972) (1959). Paris : Editions sociales, 1976 (1959)
20. Lewin, K.: Group Decision and Social Change (1952). In: Swanson, Newcomb and Hartley (eds.) Readings in Social Psychology, Holt (1952)
21. Lomov, B.F.: Man and Technology (Outlines in Engineering Psychology) (1963). Tr. JPRS 22300. US Dept of Commerce, Washington DC. December 16, 1963. p. 307 (1963)
22. Minsky, M.: La société de l'esprit (1985). Paris: Interéditions, 1988 (1985)
23. Moscovici, S.: La psychanalyse son image et son public (1961). P.U.F, Paris (1961)
24. Norman, D.A.: The Psychology of Everyday Things. Basic Books, New York (1988)
25. Nosulenko, V., Samoylenko, E.: Analyse de l'activité et numérisation. Paris: Maison des Sciences de l'Homme et EDF R&D, 2006. p. 80 (2006)
26. Rabardel, P.: Les hommes & les technologies: approche cognitive des instruments contemporains. Armand Colin, Paris (1995)
27. Rasmussen, J.: Skills, rules, knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE Transactions on Systems, Man. and Cybernetics 13, 257–266 (1983)
28. Rasmussen, J.: The role of hierarchical knowledge representation in decision making and system management. IEEE Transactions on Systems, Man. and Cybernetics 15, 234–243 (1985)
29. Rubinshtein, S.L.: Osnovy obshchei psikhologii. Bases of general Psychology.? Izdatel'stvo AN SSSR. Izdatel'stvo Gosudarstvennoie uchebno-pedagogicheskoie izdatel'stvo Narkomprosa RSFSR (1940)

30. Rumelhart, David, E., Norman, Donald, A.: Representation in Memory. Center for Human Information Processing. University of California, San Diego. Report no. ONR 8302. UCSD Chip 116. June 1983, p. 117 (1983)
31. Shank, Roger, C., Abelson, Robert, P.: Scripts, Plans, Goals, and Understanding. Lawrence Erlbaum Associates, Hillsdale, NJ (1977)
32. Simon, H.A.: Administrative Behavior (1945). Free Press, N.Y (1945)
33. Simon, H.A., Newell, A.: Human Problem Solving. Prentice-Hall, Englewood Cliffs, New Jersey (1972)
34. Suchman, L.: Plans and situated actions. In: The problem of human-machine communication, Cambridge University Press, Cambridge (1987)
35. Theureau, J.: Le cours d'action: analyse sémio-logique, essai d'une anthropologie cognitive située. P. Lang, Bern (1992)
36. Uexküll, J.V.: Mondes animaux et monde humain (1921). In: Suivi de Théorie de la signification, Médiations, Gonthier, Paris (1965)
37. Veblen, T.: The Theory of the Leisure Class (1899). Modern Library, New York (1834)

Empirical Evidence for a Model of Operator Reaction to Alerting Systems

Steven J. Landry and Anil Divvela

School of Industrial Engineering, Purdue University
315 N. Grant St. West Lafayette,
IN 47906, USA
{slandry,adivvela}@purdue.edu

Abstract. A sample of aviation safety reports related to the Ground Proximity Warning System was pulled from the Aviation Safety Reporting System. These reports were used to evaluate a model of operator reaction to alerting systems, particularly focusing on cases of noncompliance. The model posited that operators try to define a causal path from the distal situation triggering the alert (which is generally unknown) to the proximal cues. The model implies that alert response behavior can be influenced by manipulating those factors that affect the ability to construct this mapping. Based on the examination of the reports, two refinements of the model are proposed. First, the ability to define a causal path is affected by time pressure; only those paths that can be verified are checked. Second, the paths are checked against preconceptions based on the task situation. These findings may have implications for the design of procedures and alerting systems.

Keywords: alerting systems; aviation; situation awareness; procedures.

1 Introduction

In studying accident or safety reports which involve operators reacting to alerting systems, it is frequently found that such reactions are slow, are different than expected, or are sometimes even ignored [1, 2]. This has been found regardless of the salience or importance of the alert, and regardless of the relative cost of the possible outcomes associated with ignoring or attending to the alert.

For example, there have been a number of controlled-flight-into-terrain accidents, where fully operational aircraft crash into terrain, despite alerts from the ground proximity warning system (GPWS) identifying the approaching terrain danger several minutes before the incident. The pilots acknowledged the alert but believed it was a false alarm, and so continued without executing the procedure associated with the alert.

Typically this type of noncompliance is attributed to a lack of confidence in the alert [3-7]. The explanation is that since the GPWS was known to have false alarms, pilots will sometimes ignore the alert. The remedy that follows from such a description is to reduce false alarms. The thesis of this work is that this explanation, and in particular the design recommendations stemming from it, is insufficient. We

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 522–530, 2007.

believe that operators have a compulsion to identify a particular logical path from the undesirable distal situation (the approaching terrain) to the proximal stimulus (the GPWS alert) in order to confirm the presence of the undesirable distal situation prior to initiating the actions called for by the alert-reaction procedure.

In the paper we will discuss how this compulsion is manifest as information-seeking behavior which occurs after the alert and prior to initiation of the alert-reaction procedure. It is argued in this paper that this information-seeking behavior is a necessary prerequisite for failures to comply with alert-response procedures. The reason for this is that the failures to comply seem to come mainly from support operators find for faulty assumptions about the situation which is causing the alert. For example, if a pilot assumes that a GPWS warning is false (which is likely if the pilot does not trust the system), then seeing no dangerous terrain identified on the radar display is support for that conclusion. That pilot is then likely to ignore the GPWS warning, despite the overwhelming cost difference between an ignored true alert and a successfully detected false alarm.

The importance of this work is that, if true, this model of alerting system reaction would prescribe different design criteria for alerting systems, affect the design of procedures, and possibly suggest different training for operators of the system. Instead of the sole focus being on reducing false alarms, significant attention would need to be paid to the information that operators may seek to determine the credibility of the alert.

In general, pilots were found to search for information in order to confirm a particular, apparently pre-conceived, mapping within the model, and reacted according to the information that corresponded with that mapping. Pilots sought to very quickly identify certain information in order to come to a decision as to whether the alert was valid, even when such information only partially correlated with the state the pilot was trying to infer.

1.1 Description of the Model

A generic model of an alerting system-operator system is shown in Figure 1. Initially, the model was based solely on the idea that operators will search amongst the space of possible reasons for the alert to have occurred, trying to identify the logical path which triggered the proximal stimulus (i.e. the alert, typically aural and/or visual). As will be discussed, additional details regarding the effect of time pressure and preconceptions were added after examining the safety reports.

In Figure 1, there is some hazardous situation that the alerting system is attempting to identify based on alert criteria, some of which are probabilistically associated with the hazardous situation. Combinations of these criteria trigger an alert, which belongs to one of a number of condition sets. These same sets of conditions, when added to some other condition, can lead to false alarms. These other conditions are likely not able to be filtered out (for example, they may be malfunctions) or may be unforeseen by the designers of the alerting system.

The operator has access only to proximal cues which correlate with the distal situation or the alert criteria (neither of which are directly available to the operator). On the basis of these proximal cues the operator must identify which condition set has triggered the alert.

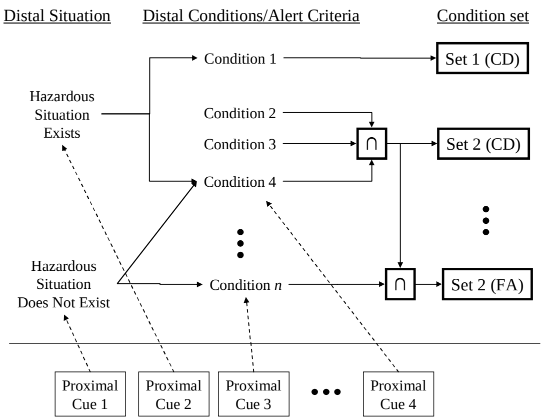


Fig. 1. Alerting system-operator model

There are several aspects of uncertainty which lead to the indeterminacy of alerts. These include that the correlation of the proximal cues to the distal situation (or distal conditions/alert criteria) is typically less than one, as is the correlation of alert criteria to distal situation.

The model posits that operators search among the proximal cues for information that will identify the condition set that has been triggered. This information then allows them to make a decision about a proper course of action. As a consequence, operator performance depends on the ability to make a proper mapping from distal situation to condition set.

1.2 Enhanced GPWS: Description

The Enhanced GPWS (EGPWS) is an enhancement to an earlier system – the GPWS. The GPWS was designed, in response to a number of "controlled flight into terrain" accidents, to detect unsafe closure to terrain and warn the pilot. It uses configuration information, descent rate, and a radar altimeter (which measures height above the ground). The GPWS alerts the pilot using both visual and audible alerts; the audible alert was a synthesized voice which commands "PULL UP."

The EPGWS typically uses a terrain database, a terrain display, and a set of additional alerts. It is actually built as an addition to GPWS, and the two systems operate independently. However, in a practical sense the two systems appear as one system, and for the remainder of the paper the term EGPWS will be used to refer to the combination of the two systems.

The EGPWS system is extremely complex, including its alert criteria. One representative system, Honeywell's MK VIII EPGWS [8], includes nine alert modes, some of which more than one possible annunciation. Each of these alert modes activate under a set of criteria. The criteria and annunciations shown in Table 1 are actually a simplified version of the actual criteria. For example, the alerts usually change from an initial annunciation to a more forceful annunciation if the conditions do not change for a certain period of time, or if different criteria (associated with the predecessor GPWS system) are met. There are also different criteria for different

Table 1. Modes of the Honeywell Mk VIII EGPWS

Mode	Criteria	Initial Aural Alert Annunciation
1	Aircraft penetrates nonlinear altitude/closure rate envelope (different for turboprops and jets), adjusted for valid ILS glideslope indication, modified by "steep approach" selectable option and "flap override" option.	"SINKRATE, SINKRATE"
2A	Flaps not in landing configuration, aircraft not on glideslope centerline, aircraft penetrates nonlinear altitude/closure rate envelope (dependent also on aircraft speed), modified by terrain alerting and display function option.	"TERRAIN, TERRAIN"
2B	Flaps in landing configuration, glideslope and localizer deviation less than two dots, aircraft within 5 nautical miles and 3500 feet of destination airport, terrain alerting and display function option selected, aircraft penetrates nonlinear altitude/closure rate envelope.	"TERRAIN, TERRAIN"
3	Aircraft below 245 feet above ground level, aircraft penetrates nonlinear altitude/altitude loss envelope, modified by alternate mode 4B options, modified by flap override option.	"DON'T SINK, DON'T SINK"
4A	Gear not in landing configuration, aircraft below 500 feet above ground level, airspeed below 178 knots, aircraft penetrates nonlinear airspeed/terrain clearance envelope.	"TOO LOW GEAR"
4B	Gear in landing configuration, flaps not in landing configuration, aircraft below 170 feet above ground level, airspeed below 150 knots, aircraft penetrates nonlinear airspeed/altitude envelope, modified by aircraft type (turboprop or jet).	"TOO LOW FLAPS"
4C	Either flaps or gear not in landing configuration, aircraft below 245 feet AGL, aircraft penetrates nonlinear altitude/terrain clearance envelope, modified by aircraft type, modified by alternate mode 4B options.	"TOO LOW TERRAIN"
5	Aircraft below 1000 feet above ground level but above 150 feet above ground level, aircraft 1.3 dots or greater below glideslope, aircraft within 2 dots of localizer, landing gear and flaps are selected, glideslope cancel not active, front course approach is used (determined by the automation), aircraft penetrates nonlinear glideslope deviation/altitude envelope.	"GLIDESLOPE"
6	Reaching predefined altitudes (preset at installation)	Dependent on setup, e.g. "FIVE HUNDRED"

aircraft, or even different models of the same aircraft. The criteria shown are just to provide an indication of the difficulty of actually understanding the specific conditions that caused the alert to sound.

1.3 EGPWS: Model

We constructed a mapping of distal situation to condition set for the EGPWS-pilot system, where condition sets were, in part, the alert modes in Table 1. This mapping is too large to reproduce here, but a portion of that mapping is shown in Figure 2. The actual alert criteria can be broken down into more detail, but that detail does not necessarily add explanatory power to the model.

In the model the EGPWS is trying to ascertain whether terrain danger exists. The aircraft penetrating the envelopes described in Table 1 and Figure 2 is correlated with such terrain danger, along with other criteria such as the configuration of the aircraft.

The pilot is unable to determine whether the aircraft has penetrated these envelopes, as there are many of them, all of which are complex. While theoretically the pilot could examine them all with reference to a manual, such a procedure would be very time-consuming. However, the pilot has easy access to a large number of perceptually available cues, such as direct visual information about terrain (if not in the weather), a terrain map display, landing gear indicators, flap indicators, vertical speed indicators, etc.. These cues correlate either with the distal situation or with

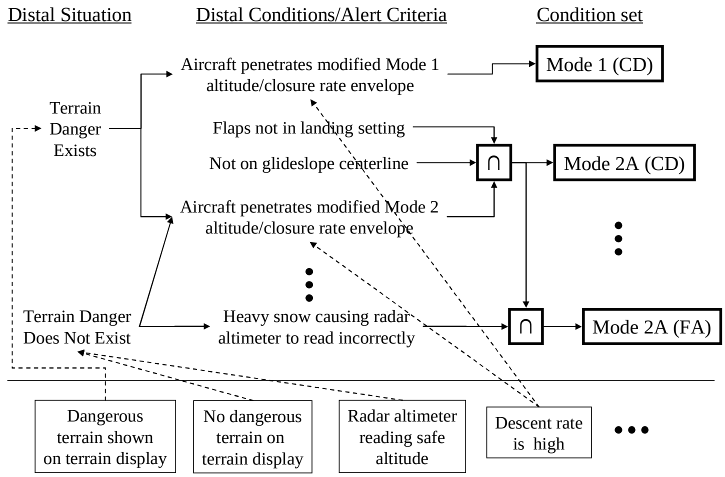


Fig. 2. Partial model of EGPWS-pilot system for alert conditions

some of the alert criteria. By checking these cues, pilots can make an informed decision about what condition set has activated, and thereby determine whether the alert was correct or false.

1.4 EGPWS: Mandated Pilot Reaction

Although a complete survey of airline policies and procedures was not completed, it is believed that the standard response to an EGPWS alert is the same as was mandated by the FAA in a regulatory circular. That circular states that "when an EGPWS CAUTION occurs, adjust the airplane flightpath until the CAUTION alert ceases" and "if an EGPWS WARNING occurs, immediately initiate and continue a climb which will provide maximum terrain clearance or similar approved vertical terrain escape maneuver, until all alerts cease [9].

Honeywell recommends that cautions (the annunciations in Table 1) be responded to by some investigation prior to initiating action, but recommends the same course of action as the FAA for warning ("PULL UP") alerts which follow cautions if conditions do not change.

2 Method

One hundred reports were pulled from the ASRS database from a search on commercial (i.e. Part 121) flights using the keyword "GPWS." Of these, 64 reports involved an actual alert. These 64 reports were reviewed in detail.

The responses of the pilots were categorized as either "immediate," "delayed," or "no response." The response was also characterized as either "correct" or "incorrect." The alert was categorized as a "false alarm" or "correct detection" whenever there was sufficient detail in the report to make a determination. For those responses

labeled “delayed” or “no response,” a description of the pilot’s actions was recorded along with as much information about the task environment as was provided.

This information was checked against the EGPWS model to see if pilots were searching for information in order to map the path from condition set to distal situation (see Figure 2).

For example, the following is the first part of the narrative text from an ASRS report (abbreviations and missing words, common in these reports, are expanded in square brackets to differentiate with the original parentheses):

> On [approach] to XNA (Northwest Arkansas Regional [Airport]), my [first officer] was flying [a] visual [approach] to [runway] 34. Level at 3000 ft, we received [an] EGPWS warning “terrain alert.” (Our company’s [procedure] is to correct for warning at night or IMC without question.) The EGPWS radio altimeter nor did the terrain display auto display on [the primary flight display] or [the multi-function display]. My [first officer] made no correction and stated “it’s nothing.” I told him, “pulling up, max [power] spoilers in” and to [climb] and correct. The warning sounded again and he started [climbing] … [10]

This report would be interpreted as having a delayed response, but a correct response The pilot, before initiating the procedure, which was supposed to be initiated “without question,” checked the radio (i.e. radar) altimeter and the terrain display. This appeared to be the only reason for the delayed reaction (at least on the pilot’s part – no information is provided on why the first officer delayed and chose to ignore the alert). The report also indicates that it occurred in visual conditions at night. The rest of the narrative did not provide sufficient detail to conclude whether the alarm was correct or false.

3 Results

The results are summarized in Table 2.

Table 2. Categorization for 64 ASRS reports on EGPWS incidents

	Immediate and correct	Delayed and correct	Immediate and incorrect	Delayed and incorrect	Totals
FA	2	1	0	3	6
FA (likely)	10	4	2	0	16
Unknown	18	3	0	1	22
CD	8	1	0	0	9
CD (likely)	10	0	1	0	11
Totals	48	9	3	4	64

On the basis of this sample, the false alarm rate is from 6/64 to 22/64 (9% - 34%), without considering the unknowns as possible false alarms. The pilot performed an immediate action 51/64 times (80%), performed a correct action 57/64 times (89%).

Of the 16 reports which showed some anomaly (either delay or incorrect response), all but one were during the approach phase. The remaining incident was during departure. Twelve of the incidents occurred either at night or in non-visual conditions.

Of the 13 reports where there was a delayed reaction, the delay occurred solely because the pilot was obtaining additional information in all but one case. In the one case where information-seeking was not the only reason for delay, the pilot flying failed to immediately acknowledge the alert for unknown reasons. However, the pilot also reported looking for additional information after the initial unexplained delay.

The information these pilots reported checking varied, although in 7 of the 13 incidents the terrain display was checked. This display is part of the EPGWS system and shows the terrain that caused the alert on a map display that can also be used for the weather radar and navigation information. Three pilots checked the radar altimeter (which shows the altitude above ground level), two pilots checked the terrain visually, and two pilots checked navigation instruments. Other information checked in only one report each were the aircraft configuration, whether the checklist had been completed, the barometric altimeter (which shows the altitude above mean sea level), the descent rate, and the glideslope (which shows the proper descent path to the runway).

4 Discussion and Future Work

First, the data should be interpreted with caution. The ASRS database is a collection of voluntary, subjective reports based on the recollection of the pilot(s). It is not a scientifically collected statistical sample, and the information contained has not been verified (as would an actual accident report). The reports are often vague about details, forcing us to interpret what actually occurred. Finally, the specific EPGWS response procedures mandated for the particular crews involved are unknown, and the specific manufacturers and models of EGPWS systems in use on these aircraft are unknown.

In addition to these limitations, only one alerting system was studied in one domain. Alerting systems vary widely in complexity, urgency, and design. In different domains reaction to alerting systems is not normally a critical function as it often is in aviation. Many alerting systems have fairly simple criteria which operators may be able to check directly. Future work will need to investigate other domains and alerting systems to see whether the model is generalizable beyond the EGPWS and beyond the aviation context. It is thought that, at least, these results apply to high workload domains in which alerting systems are safety critical.

The reports were examined to see if any information could be found that contradicted the model. No directly contradictory evidence was found. Instead, the information in these reports did seem to indicate that delays were most frequently due to some information-seeking behavior by pilots.

A few observations identified some limitations which seem to be acting on the pilots. In the original model, pilots would actively search for information supporting each possible mapping from distal situation to condition set.

This does not seem to strictly be the case. Instead, pilots appear to have a small set of information (proximal cues) that they know correlates with the distal situations or

with the distal conditions/alert criteria. They search these proximal cues based on the time they have available, which in these cases (during approach or departure) wass very short. The cues that pilots searched (terrain radar, radar altimeter, etc.) are easily available, perceptually simple, and correlate well with the operation of the EPGWS.

Moreover, the cues pilots will check are probably influenced by the specific conditions of the task at the time of the alert. For example, the finding that the configuration was checked makes sense given that these aircraft were approaching the airport. One would not expect to find that in the case where the alert occurred during departure (and it was not).

In almost all of these cases pilots were surprised by the alert. Only in one case did the pilot note that there was concern about high terrain prior to the alert. Since they were not expecting the alert, the pilots probably did not believe that the conditions warranted a warning from the EGPWS. One would then further suspect that their initial belief may be that the alert was false. Such a belief would be obtained abductively – seemingly the best explanation given the environment at the time. Subsequent information might then be sought to confirm this hypothesis.

This type of reasoning is famously prone to the fallacy of affirming the consequent, which might explain the record of controlled flight into terrain accidents in which alerts were ignored. In those cases, the logic went something like this (for example): "if conditions A, B, C are all true then the alert is correct," "condition B is not true," therefore "the alert is false." If you replace A with "approaching terrain," B with "aircraft not configured for landing," and C with "system is operating properly," then you can see the fallacy. Just because B is not true does not mean the alert is false, because there is not an "if and only if" relationship to the first statement.

If they abductively believed the alarm was false, the cues they examine may be those that could confirm a false alarm, rather than refute it (or confirm a correct detection). In these reports, much of the information correlates with both a false alarm and a correct detection, making verification of this supposition difficult. However, in describing their interpretation of the information, pilots referenced the information as if confirming their suspicion of a false alarm. For example, one pilot stated that after hearing the EGPWS warning, "I quickly [checked] to ensure that the [first officer] had indeed leveled at 2500 ft (he did), and that we were not descending (we weren't) [11]." Future work should be able to distinguish whether the information pilots seek is done to confirm false alarms or correct detections.

If the model is accurate, it will have significant implications for alerting system design, procedures, and training. For example, the current emphasis on mitigating poor responses to alerting systems is to minimize false alarms. While that certainly has an effect on the initial, abductively-obtained hypothesis about the alert, in many cases it is difficult to reduce false alarms to a level where confidence in the system is high. Therefore, in addition to minimizing false alarms, this work suggests that one may be able to attack the subsequent information-seeking behavior to better inform pilots about the validity of the alert. In addition, alerting systems may need to consider that pilots will likely delay reaction while searching for information when the alert is unexpected. Rare, unexpected alerts that require rapid responses are not likely to be successful.

Procedures and training would also need to take into account the compulsion pilots have to seek out confirmatory evidence. Instead of mandating immediate action, pilots

should be trained on what information to seek and how to draw proper conclusions from that subsequent information.

Acknowledgments. The authors wish to thank Purdue University's Summer Undergraduate Research Fellowship (SURF) program, its sponsors (the Intel Foundation, the Alliance for Graduate Education and the Professoriate, the Louis Stokes Alliance for Minority Participation, the National Science Foundation, the Fluor Foundation, and Daimler-Chrysler), and the School of Industrial Engineering for providing support for this research. We also wish to thank Amit Lagu, who served as graduate student mentor on this project.

References

1. Pritchett, A.R.: Pilot non-conformance to alerting system commands. In: 1997 IEEE International Conference on Systems, Man, and Cybernetics (Orlando, FL, 1997), Institute for Electrical and Electronics Engineers, pp. 2125–2130 (1997)
2. Proulx, G.: Why building occupants ignore fire alarms. In: Construction Technology Update No. 42, Institute for Research in Construction, National Research Council of Canada, Ottowa (2000)
3. Bliss, J.P., Gilson, R.D.: Emergency signal failure: Implications and recommendations. Ergonomics, 41 (1). 57-72
4. Bliss, J.P., Dunn, M.C.: Behavioural implications of alarm mistrust as a function of task workload. Ergonomics, 43(9), 1283–1300
5. Bliss, J.P.: Investigation of Alarm-Related Accidents and Incidents in Aviation. International Journal of Aviation Psychology, 13(3), 249–268
6. Thomas, L.C., Rantanen, E.M.: Human factors issues in implementation of advanced aviation technologies: A case of false alerts and cockpit displays of traffic information. Theoretical Issues in Ergonomics Science, 7(5) 501–523
7. Proulx, G.: Strategies for ensuring appropriate occupant response to fire alarm signals. in Construction Technology Update No. 43, Institute for Research in Construction, National Research Council of Canada, Ottowa (2000)
8. Honeywell International Inc. MK VI and MKVIII: Enhanced Ground Proximity Warning System (EGPWS) Pilot's Guide, Morristown, NJ (2004)
9. Federal Aviation Administration. Enhanced Ground Proximity Warning System, Department of Transportation, Washington, D.C (1996)
10. Anonymous. Aviation Safety Reporting System Database Report, ACN 581622, National Aeronautics and Space Administration, Washington, D.C (2003)
11. Anonymous. Aviation Safety Reporting System Database Report, ACN 662377, National Aeronautics and Space Administration, Washington, D.C (2005)

“Investigating the Way National Grid Controllers Visualize the Electricity Transmission Grid Using a Neuro-Linguistic Programming (NLP) Approach”

P. Lazanas

University of Johannesburg
37 Nind St. Doornfontein 2094, South Africa
plazanas@uj.ac.za

Abstract. Highly skilled South African experts are responsible for controlling the voltage levels of the electrical transmission grid. This paper reports on the research methodology, which was used to identify the mental structure of their expertise. The research was multidisciplinary in nature, including, Power System Engineering, Industrial and Cognitive Psychology, and Neuro-Linguistic Programming (NLP).Of these, the NLP elicitation techniques used form the main body of this paper. NLP represents a body of knowledge with which human experience can be investigated and altered, by adopting a “whole-person” approach. Eight controllers of varied competence were interviewed. The main findings were that the expert controllers use sophisticated mental strategies that allow them to simplify the overwhelming quantity of data presented to them. These results will be discussed in the full paper together with the applicability of the NLP elicitation method.

Keywords: Mental Models, Naturalistic Decision Making, Neuro-Linguistic Programming (NLP), Human Factors, Visualization.

1 Introduction

The South African electrical power grid is one of the largest in the world. It is controlled from the national utility’s (ESKOM) control room, near Johannesburg.

The ESKOM National Control Centre (NCC) comprises a circular control room in which a number of staff interact with a wide variety of computer displays and consoles. These staff members, or controllers, are responsible for the coordination of supply and demand of electrical energy, from the various centers of generation to the different load centers around South Africa. Power is transmitted via the national grid of high voltage power lines, sub-stations and associated equipment. The ESKOM network is linked directly or indirectly to other African countries, this interconnected network is termed the Southern African Power Pool (SAPP) [1].

The aim of this paper is to shed some light on what makes these network controllers the experts that they are as very little, if anything is known about the mental processes which they employ in their work [25, 26, and 28].

Work on the Voltage Control desk demands the most skill, as voltage control is both complex and dynamic. In order to carry out their duties the controllers need to

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 531–540, 2007.

have access to a variety of technical information. This includes the available capacity of generation stations, the level of power demand by load centers and its fluctuations, voltage levels at all national grid buses, reactive power flow, weather conditions and other environmental factors as well as the state of "health" of the grid. All this information is presented in a variety of ways. There are computerized displays of single line diagrams, numerical values of critical variables and various alarms, as well as information conveyed in writing, over telephones and by personal interaction. This complex real-time flow of information has to be interpreted by the controllers. They then make complex decisions to ensure the safe and continuous operation of the whole network, within predetermined parameters.

Rosalind Williams, technology director at the Massachusetts Institute of Technology, is quoted in an article published in 2003 [2] as saying, "Engineering has evolved into an open-ended Profession of Everything". A multidisciplinary approach was used to investigate the mental world of the grid controllers.

The research described in this paper is focused on understanding the mental models and strategies used by controllers in their interpretation of the network data using an NLP elicitation approach. The lack of knowledge about how controllers use mental models and strategies became evident from preliminary interviews with senior management at the NCC. A number of difficulties arose due to this lack of knowledge. These difficulties relate to: Knowledge transfer; Training; and Accepting "new" computer display formats [1]. Symptoms of these difficulties manifest in a number of contexts. These include skills transfer and training, staff upgrading, integration between personnel and computer displays, and information gathering during investigations into operating errors. The scope of reporting on the findings of the research presented in this paper has been limited, due to space constraints, to the NLP methodology used. The full findings can be viewed in [1].

For the purpose of this paper the research question asked was:

- What Neuro-Linguistic Programming (NLP) methodology can be applied to find out how the experienced voltage desk controllers use mental representations, strategies, and visualizations of the network in order to do their work?

2 Methodology

Preliminary discussions were conducted with NCC management and staff in order to get approval for the research and to establish needs. This was a complex process in an environment where interactions take place on many social and technological levels.

A literature review was conducted to establish current trends in the fields of:
Power engineering, to identify the current practices and considerations in the field of Power Grid Control Visualization [3 to 9].

Industrial psychology, to identify elicitation methodologies. The method selected was Verbal Protocol Analysis (VPA). This method has been used in previously conducted studies, in similar South African environments [10 to 15].

Cognitive Psychology, to establish a valid theory of mental processing [16, 17, and 18]. An accepted theory of how the mind works was identified, which posits that unconscious

processing of information is at the heart of complex decision making [25, 26] and visualization.

Neuro-Linguistic Programming (NLP), to identify techniques that reveal the underlying representations, mental models and visualizations of the participants [19 to 23, 29].

From the literature review it became clear that an investigation into the mental processing used by Power Grid controllers using this approach had never been reported on in published literature. The methodology for conducting the interviews with the controllers was formulated [1]:

Seven controllers at the NCC were selected. Of these, two are experienced and skilled Senior Advisors (SA), two are Experienced Voltage Controllers (EVC) and two are New Voltage Controllers (NVC). The seventh was a Novice Female Controller (NFC) that had undergone training, but had not yet been authorized to work on the voltage desk. She was included to examine differences in perspective. An eighth person was included in the interview list. He is a Man Machine Interface Designer (MMID) but does not do shift-work or work in the control room itself. He was included in order to highlight any differences between control room staff and interface designers.

Each controller was given the same task to perform on the ESKOM grid simulator. This task consisted of a single 400 kV line trip, followed by a second 400 kV line trip within a few seconds, in the same area.

The controllers' performance on the simulator was video recorded. They were asked to verbally comment on what they were doing while doing it, as per VPA (Verbal Protocol Analysis) requirements.

The video recording was then played back to them and they commented further on their activities (Retrospective VPA).

A series of questions were asked to elicit the underlying mental representations that made their decisions possible and reveal their mental models. *It is during this phase that the NLP elicitation methodology was applied. The interviews took one whole day per person to cover the few minutes of video recorded simulator tasks.*

The video recordings of their performance on the simulator, as well as the video recordings of the interviews were analysed to extract the material that represents the findings of this research. The findings from the interviews were summarized and the relevant controller strategies and visualizations were compiled.

The emphasis of this paper is on the NLP elicitation methodology used, this forms the main body of the paper. The results will only reflect a sample from the findings.

3 The NLP Elicitation Methodology Used

The various aspects of the NLP human expertise modelling process were identified and are summarized below. Of particular significance was the use of the "Metamodel" and "Sub-Modality" elicitation techniques.

3.1 The NLP Modeling Process

As identified from the various references cited [19 to 23], the stages of modeling human expertise can be summarized as follows:

- **Elicit** (relevant data, and gather information from expert / s);
- **Formalize** (structure / simplify the model of expertise);
- **Transfer** (the model to non-experts as per individual requirements).

The experts being modeled were not fully aware of the mental representations / structures of what it is that they do. When asked, their typical response was "It's experience" or simply "I don't know". It was therefore the task of the modeler to move them into a *state* from which they could report on their experience by means of the NLP based techniques listed below. During the elicitation process it was necessary to:

Listen to the words that the subject used; *Observe* the subject's behavior; and
Identify the underlying intentions or implied mental actions of the subject.

In preparation for elicitation it was necessary to consider the following: Am I modeling behavior, internal representations or both? Review the presuppositions that have been identified during the preliminary interviews and ensure that everyone involved understands what will take place and how. The methods of elicitation used were: One on one interviews; Observation of the performance of tasks, with or without commentary; Video recording; Audio recording; Group elicitation; Mixed.

How will the data be formalized into the final model, by what means and on what level of detail? The level of elicited detail of the structure of the controllers' mental imagery was quite high.

How will the model be transferred after it has been formalized and to whom?
In preparation for elicitation, the following requirements were set for myself:

Be clear about what it is that you want to elicit, yet *flexible* enough to allow other important issues to come through. Always allow space for extra *experiential information but discourage rationalizations and long "explanations"*. Look, observe, listen, sense, and enquire about what is not explicitly offered by the expert. Avoid asking "why?", rather ask "what, where, how, when, etc." questions.

Establish a way by which you can maintain your focus on the other person and not drift into thoughts of your own. In this respect video recording is a must. Establish ways by which you can maintain your energy level during the interview process.

Establish ways to either associate or disassociate from the subject to gain varied perspectives.

Ensure that you meet the individual you are interviewing at their model of the world. Always ask questions to verify what is been said, *never suggest to the other person what you think is or should be going on*.

Be familiar with the jargon of the field in which the person you are interviewing works. In this respect the multidisciplinary approach really worked well.

During the interview process it was possible to go from macro aspects (I am a controller in a defined environment); to Micro aspects (*Sub-modalities* see below.)

In each case notice the internal and external, conscious and unconscious, overt and covert aspects of the subject's experience and enquire about what you observe. To show a subject video footage of them performing a task was an excellent way to start the elicitation process.

Choose the format (mind map, lists etc.) and the material you will use (note pad, flip chart etc.) to log the responses in the following dimensions of the modelling subject. The terms below refer to what is known in NLP as *"Logical Levels"* of

organization of experience or *"Neurological Levels of Organization"* [22]. A similar view is also presented in Pinker's, Anderson's, and Shalif's work. They however do not use the term logical levels, instead they speak of the way the brain is structured in expert "modules", which interact with each other in neurological hierarchies of organization. The deeper layers of the information presented by the controllers were "probed" with this framework. The idea is that behaviors are possible because of abilities; which are in turn activated by beliefs and motivated by values or the desire to experience certain states of being; these states as a collective define an individual's identity and that identity is then considered as part of a greater context or whole. This is by no means a linear process. Starting from the top (higher complexity), these *Logical Levels* are:

Beyond: This level represents an individual's "connection" with larger wholes or structures like work group, family, nation, the universe, God, etc.

Identity: His / her sense of individual self. This refers to the controllers' sense of identity within the control room environment. "Who are you with respect to the network?" was the question asked. The controllers' sense of identity outside the control room was also elicited; in some cases the two were quite different.

Values: A value is defined here as "that which has emotional impact", with respect to their work and other experience. What they fear was also considered. Shalif [18] states that "we are motivated by our need to achieve emotional states rather than the results of our actions". We use whatever skills we have to achieve results so that we can experience the emotional states they create. "How will you *feel* and in what way, when you have what you want?" is a typical question asked in order to elicit values. The physiological characteristics of their words were noted, i.e. "heat in the chest", a "lump in the throat", as part of the sub-modality elicitations of these words.

Beliefs: What were elicited were the beliefs which the controllers hold about the network, themselves, their ability to control it, and how their work environment will affect them in the long run. "Why?" was the preferred question to elicit beliefs.

Abilities: The underlying abilities, behind the controllers' behaviours, were identified first from what they did on video, followed by their subconscious abilities.

Behaviors: These were elicited from the video footage while subjects were performing the allocated task. They are the "what" of that which was observed.

Environment: The NCC was examined as a working environment.

Together with the categories above the following were investigated:

Sorting Styles: These are the filters (sort through = filter) through which information passes. Controllers were asked to indicate what it is that they sort for i.e.:

Information, Others, Activities, Inspiration, Things, Past, Present, Future, Relationship, Associate, Dissociate, Approach, Avoidance, Match, Mismatch, Places, Power, Positive, Negative, Exterior, Interior, Feeling, Centeredness, Other…

Metaphors: By the use of metaphors we can incorporate the technique of "this is like…" in order to use qualities and complex correlations implicit in the metaphor and "map them over" to the idea or topic at hand. For instance to say that a man is like a "bear" implies a lot of complex ideas associated with the word "bear" that go much further than saying that he is "big". "What is the network like?" "How do you relate to the network (externally and internally)?" were typical questions asked.

Metamodel: This represents a method of identifying and overcoming the linguistic violations listed below. It involved asking specific questions in such a way as to

reveal the hidden, deeper level, meaning of the words that the controllers used. By using this technique during the interviews the finer detail of the controllers' inner abilities, beliefs, values etc. were identified. Below is a list of linguistic violation terms as defined in the literature [23, 24].

Deletion: Important information which is left out. This limits thought and action. I.e. Simple / Comparative Deletion; Unspecified Verb; Unspecified referential index; Judgments; Comparisons.

Generalization: One example of a class is taken to represent the class, in ways that narrow possibilities. I.e. Modal operators of necessity; Modal operators of possibility; Universals.

Distortions: Information is twisted. I.e. Nominalizations; Mind reading; Cause and effect; Complex equivalences; Presuppositions.

The other very important aspect of the interview process is the *modality and sub-modality elicitation* [19 to 24].

Modalities are the five senses through which we process all information, as well as represent it to others through the medium of words. The five primary senses (Visual, Auditory, Kinesthetic, Olfactory and Gustatory), were considered even though there are others, i.e. balance etc. For the purpose of this research only the visual, auditory and kinesthetic senses were examined, as these are mostly relevant in the NCC environment. In this section the sensory structure of the meaning of words was elicited. In other words, if a controller mentioned the word "grid" it was required of him / her to describe in what way exactly he experienced the "grid". Did he see it as an image, hear it as a sound, feel it as a feeling, or a combination of these? Shifts in modalities were noted as the person went further into their deep-structure experience.

The *Sub-modality* aspect of the sensory structure of experience was the last layer that was elicited. An individual's expertise is structured in terms of experiential components that operate very much like computer software [16, 17]. For instance, using the example of the word "grid"; if the experience of the word "grid" was visual it was determined whether the image was in color or black and white, how far from the viewer it appeared (its sub-modalities) etc. In this way the controllers' mental worlds were explored.

Here are some Sub-modality examples: Black, White, Bright Dim, 3D, Flat, Distant, Close, Movie, Still, Movement rate, Tone, Tempo, Pitch, Sound direction, Volume, Constant, Intermittent, Location, Pressure, Heat, Heavy, Light, etc.

For each modality the sub modality structure of each individual's experience was examined (this was a slow process that required a lot of patience). Having sought permission, layer after layer of the structural components of the mental representations used by the controllers was uncovered.

Strategies. These are the *sequences* of Visual, Auditory, Kinesthetic, Olfactory, Gustatory elements of the "software" which makes performance possible, as well as the Test, Operate, Test, Exit (T.O.T.E.) [24] patterns pattern which people use to initiate and terminate sequences of behavior. In this section the *sequencing* aspect of the structure of experience or expertise was identified. For instance, a controller "saw" a mental image of the whole of the ESKOM grid "then zoomed" into a particular section etc.

The meaning of verbal communication largely lies in intonation and body posture [23]. As a result special care was taken to notice subtle, unconscious movements, for clues to the underlying states of the words used by the controllers. Often, what was being conveyed by a person's physiology was truer than what was being revealed through their words. In this respect the video material was invaluable. Incongruence between what was being said and the speaker's physiology was explored as far as possible. The following were considered:

Mannerisms (patterns), as well as anything that stood out by overdoing or not doing; use of the left or right side of the body; hand and finger movements; head tilt and eye movements and what they stood for (visual, auditory, kinesthetic, positive / negative, past, present, future); feet; mouth; hidden vs. exposed parts; conscious vs. unconscious movements; direction (facing, opposite, with, etc.); etc.

Physiology. Examples of physiological observation aspects that were considered are: Breathing, Location of tension, Muscles, Eye Blink, Gestures, Posture,Head nods, Head location, Mouth movement, Hand movements, etc.

In NLP terms, emotions are very important aspects of experience. As mentioned above they are considered under the broad "kinesthetic" modality and sub-modalities.

The ability to have "perspective" on a situation depends on how a person is "positioned". In general, when a person is fully associated with an experience they are able to feel and act; when they are dissociated they are more able to evaluate and make comparisons. Further, the ability to empathize with others, to see / feel things from the other's point of view, is also an important ability when it comes to teamwork.

Perceptual Positions. In relating their experience the controllers were in one of the following perceptual positions.

First: Describing events through their own eyes and feeling their own feelings.

Second: Describing events through the eyes of another and relating to their emotions.

Third: Describing events from a distance, looking on the events described with detachment.

The perception of time is subjective, both with respect to "real" time as well as whether a person is focused on the past, present or the future. The flexibility to be able to focus on any one of these time domains, at will, is part of what makes up an expert controller. This is because they have to remember / distinguish past events, clearly sort out current information, and anticipate events and the consequences of their decisions. The time orientation of the controllers was noted as *Past, Present, or Future*. Hand locators or other expressions and gestures for past, present, and future were noted.

Other Evidence. Other evidence such as: drawings; environmental factors; jargon; symbols; food stuff; traditions; rituals; rules; technical knowledge; theories etc.; was collected and added to the overall set of material that was used to make explicit the controllers' inner worlds.

4 Findings

By reviewing the video material and extracting the relevant information a matrix which contains sixty observations or statements was created. In this matrix what the participants said during their interviews is listed on the vertical axis and the eight participants are placed on the horizontal axis. X marks were inserted where comments came from a particular participant [1].

What follows is a condensed version of the responses of one of the experts (SA).

"I have a strategy to do a "gap" analysis to see if voltage levels are within predetermined values used as references. These references are self-created."

"I have a mental template that has a low and a high value on it; I then use this mental template to scan the voltage level information (in text form) on the monitors in front of me. I do not remember / register the actual voltage values that appear on the monitors if they fall within my template references."

"I experience intense feelings related to how big the difference is between my (mental) reference and the actual voltage value on the monitors. When my voltage limits are approached I feel "orange" alarm feelings (pressure in my gut) when they are exceeded I feel "red" alarm (butterflies in my gut) feelings. I have three categories of pre-determined operating sequences, normal, abnormal, and critical. When an abnormal or urgent feeling comes up I get pictures in my mind of what to do."

"I keep working until voltage values are as per the desirable reference."

"I scan key reference voltage points in the regions as per level of priority."

"I make decisions within seconds during emergencies."

"I use simple block diagrams in my mind to isolate an area on which I work." *This was termed "rubber banding" and was the most significant difference between the experienced and the inexperienced groups of controllers.*

"I compare what I see on the monitors with the above simplified block diagram in my mind. This mental image is simple (a rectangle with some lines coming out of it in black and white) and contains about six key elements which represent a substation or region and the main power lines to and from it."

"I use it to direct me as to which sector or equipment to work with next. Using this mental image I visualize the consequences of my operations on other areas in the network. I mentally drop information that I don't use."

"At other times I can visualize the whole or parts of the network as required, the image opens up in front of me like a sphere in color. I can then zoom in or out."

"I (mentally) "step back" and see the big picture in my mind regularly."

"My control sequences vary, depending on the available time to respond."

"When there is a fault I ask what caused it and recall if it happened before."

The information above represents a brief sample of data from the hours of video recording obtained (over forty hours of video).

5 Conclusions

The main conclusion has been that the expert controllers display great mental visual dexterity. That is both in terms of the complexity of their imagery, as well as their ability to rapidly mentally simplify the overwhelming plethora of information during

emergency conditions and work with simple mental diagrams. In this way they maintain a sense of what is important when as many as a few thousand alarms are been displayed during contingencies.

The information gathered from the novices (not shown above) showed that they did not have complex mental imagery (one person) or did not know how to zoom in and out of their mental maps. This was particularly severe during emergencies where their ability to mentally work with the network simply failed. During non emergency times this was not a problem.

Further, in comparison with the findings from other researchers [10 to 15], the depth of information retrieved is perhaps the main difference. Using the NLP methodology a great deal of what informs and supports the expert's ability was revealed in a way that suitable training can be designed to upgrade the novices. Verbal protocol alone does not achieve this result. A word of caution is that only a properly trained person should undertake this kind of work so as to avoid transference or "installing" of information in the mind of the subject (false memories etc.).

As far as future research goes the process of transferring the mental skills identified as well as considering the dynamics of the whole NCC environment as one complex system will be taken up [27].

Acknowledgements. I wish to thank Prof A Thatcher, Prof. B Dwolatzky from Wits University in Johannesburg, as well as all the ESKOM staff who made this work possible.

References

1. Lazanas, P.: An Investigation into the Visualization of the Transmission Network, by National Grid Controllers', MSc Dissertation University of the Witwatersrand, Johannesburg (2005)
2. Williams, R.: Education for the Profession Formerly Known as Engineering, For Engineering Educators magazine, CREE, University of Cape Town, Vol 7(1) (June 2003)
3. Mahadev, P.M., Christie, R.D.: Envisioning Power System Data: Concepts and a Prototype System State Representation, IEEE Transactions on Power Systems, vol 8(3) (August (1992) 0885-8950/93/1992 IEEE.
4. Mahadev, P.M., Christie, R.D.: Envisioning Power System Data: Vulnerability and Severity Representations for Static Security Assessment, IEEE Transactions on Power Systems vol 9(4) (November 1994) 0885-8950/94/1994 IEEE
5. Overbye T.J.: New Techniques for Power System Visualization Under Deregulation, IEEE, 0-7803-5935-6/00/2000 IEEE.
6. Overbye, T.J., Weber, J.D.: New Methods for the visualization of power System Data, IEEE Computer Society, 0-7695-0804-9/2000 IEEE
7. Overbye, T.J., Weber, J.D.: Visualization of Power System Data, In: Proceedings of the 33rd Hawaii International Conference on System Sciences (2000) 0-7695-0493-0/00/2000 IEEE.
8. Christie, R.D., Mahadev, P.M.: Visualizing Power System Data, EPRI TR-102984 Project 8010-25 Final Report (April 1994)
9. Overbye, T.J.: Visualization of Power Systems, Final report. PSERC Publication 02-36 (2002)

10. Thatcher, A.J.: A Cognitive Task Analysis in a Complex Process Control Environment, MSc Dissertation University of the Witwatersrand, Johannesburg (1995)
11. Rasmussen, J.: Mental Models and the Control of Action in Complex Environments, Mental models and Human-Computer Interaction 1. Science Publishers B.V (north Holland) (1990), Amsterdam (1990)
12. Acerman, D., Tauber, M.I.: Mental Models and Human Computer Interaction 1 & 2. Elsevier Science Publishers, Amsterdam (1990)
13. Bainbridge, L.: Cognitive Processes in Complex Tasks, Introduction and Discussion. Ergonomics 36(11), 1273–1279 (1993)
14. Bainbridge, L.: Analysis of Verbal Protocols from a Process Control Task, Mental models and Human-Computer Interaction 1. Elsevier Science Publishers B.V (north Holland), Amsterdam (1990)
15. Bainbridge, L., Beishon, R.J., Hemming, J.H., Splaine, M.: A Study of Real-time Human Decision-making using a Plant Simulator, O.R.Q. Special Conference Issue (undated, pre 1994) (1994)
16. Pinker, S.: How the Mind Works, Penguin books (1999)
17. Anderson, J.R.: Cognitive Psychology and its Implications, 4th edn. W H Freeman & Co, New York (2002)
18. Shalif, I.: The Emotions and the Dimensions of Discrimination Among Them in Daily Life. A study towards a Ph.D. dissertation, Bar-Ilan University (1987)
19. Grinder, J., Bandler, R.: The structure of Magic Vol. I & II, Science and Behavior Books Inc. (p 39-56 Vol I, p 3-26 Vol II.) (1975)
20. Bandler, R., Macdonald, W.: An Insiders Guide to Sub-Modalities, Meta Publications (1988)
21. Bandler, R.: Magic in Action. Meta Publications (1992)
22. Dilts, R., Grinder, J., Bandler, R., De Lozier, J.: Neuro-Linguistic Programming (The Study of the Structure of Subjective Experience) vol. I, Metapublications (1980)
23. O'Connor, J.: NLP Workbook, Element (2001)
24. Dilts, R.: Modeling With NLP, Meta Publications (1998)
25. Klein, G.: Sources of Power: How People Make Decisions. MIT Press, Cambridge (1999)
26. Zsambok, C.E., Klein, G.: Naturalistic Decision Making. Lawrence Erlbaum Associates, Mahwah, NJ (1997)
27. Jackson, M.C.: Systems Thinking. Wiley, Chichester (2003)
28. Senge, P.M.: The Fifth Discipline, Currency Doubleday (1994)
29. Korzybski, A.: Science and Sanity 5th Ed., Institute of General Semantics (2000)

Diagnosticity of Cardiac Modes of Autonomic Control Elicited by Simulated Driving and Verbal Working Memory Dual-Tasks

John K. Lenneman and Richard W. Backs

Central Michigan University
Mount Pleasant, MI USA

Abstract. The present study investigated the diagnosticity of autonomic space for heart rate by elucidating psychological-physiological mappings in a dual-task driving simulation. Based on the results of previous studies, we predicted patterns of cardiac autonomic activity that would be elicited by dual-task driving with a verbal working memory side-task. The results generally supported the predictions made, and indicate that cardiac autonomic modes of control can be diagnostic with regard to the attentional processing resources used during task performance in the driving domain.

Keywords: Autonomic space, cardiovascular psychophysiology, mental workload, simulated driving, dual-task.

1 Introduction

The task of driving an automobile is becoming increasingly complex because of the introduction of new technologies such as navigation aids, embedded cellular phones, and on-board hard drives for audio entertainment. It has been estimated that 13.2% of all police-reported accidents specifically involve distraction [1]. Although the estimated percentage of accidents caused by driver distraction varies, all estimates suggest that driver distraction is a significant factor in causing accidents [2]. Further, the proliferation of small portable devices that can be easily brought into the vehicle may increase in the role of driver distraction in vehicle accidents.

Over the years, the amount of research focused on driver distraction has increased substantially [3]. In general, the research has shown that the addition of side tasks to the primary task of driving can have detrimental effects on driving performance. Numerous performance, subjective, and physiological measures of attentional resources have been used to assess the effects of divided attention in the driving domain. It is desirable that these measures determine when a task is imposing demands on a driver's attentional resources (i.e., *sensitivity*) and provide insights as to the specific psychological processes (e.g., motor processing, central/executive processing) imposed by the task demand (i.e., *diagnosticity*).

Some physiological measures are considered to be more diagnostic than traditional driving performance measures [4-5]. A physiological measure is diagnostic to the extent that it indexes specific psychological processes; that is, the extent to which the

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 541–550, 2007.

measure exhibits a one-to-one psychological-physiological mapping. The next sections specify our approach using the pattern of autonomic nervous system (ANS) activity that determines heart rate to increase the sensitivity and diagnosticity of cardiac measures of attention in the driving domain.

1.1 The Doctrine of Autonomic Space

The heart is dually-innervated by the sympathetic (SNS) and parasympathetic (PNS) branches of the ANS. SNS activation causes an increase in heart rate, whereas PNS activation causes a decrease in heart rate. In the classic model of ANS function, activity of these branches was thought to be reciprocal: activation of one branch was always coupled with withdrawal of the other branch [6]. The "Doctrine of Autonomic Space" posits that ANS activity is multi-dimensionally determined instead of only reciprocally-coupled [7-9]. In addition to the reciprocally-coupled modes, the SNS and PNS branches can be non-reciprocally-coupled (co-activation or co-inhibition of each branches), or even uncoupled (change in one branch without change of the other). Thus, instead of a change in heart rate being due to activation of one branch and withdrawal of the other, eight modes of autonomic control exist (see Fig. 1). Backs [10] has suggested that the advent of autonomic space increased the diagnosticity of cardiac measures because the control modes for heart rate are thought to map to the attentional resources defined by Wickens's [11] multiple resource theory.

Non-invasive measures of underlying SNS and PNS neural activity are needed to determine the mode of autonomic control responsible for the observed heart rate. In the current study, pre-ejection period (PEP) and respiratory sinus arrhythmia (RSA) were the measures used to assess the modes of autonomic control for the heart. Previous research using pharmacological blockades has shown that PEP and RSA can serve as valid measures of cardiac SNS and PNS neurogenic activity respectively [12]. Increases in PEP reflect decreases in SNS activity and decreases in PEP reflect increases in SNS activity; whereas, increases in RSA reflect increases in PNS activity and decreases in RSA reflect decreases in PNS activity.

1.2 Multiple Resources

Wickens's [11] multiple resource theory proposed that the amount of attentional resources available for task performance is limited and that resource structure can be described by four different dichotomies: two stages of processing (early and late); two modalities of perception (auditory and visual); two codes of processing (spatial and verbal); and two visual channels (focal and ambient). According to Wickens, the attentional resources required for early stages of processing such as perceptual and central processing stages (e.g., stimulus recognition, encoding) are the same, and that they are distinct from the attentional resources required for later stages of processing such as the generation and execution of a motor response (e.g., voice response, pressing a button). Similarly, he suggests that the attentional resources required to process information presented in the visual modality are at least somewhat distinct from those required to process information presented in the auditory modality. Further, he suggests that the resources to process information of a spatial nature are

distinct from those used to process information of a verbal nature. Finally, Wickens suggests that the resources available to process information using focal vision (required for fine detail and pattern recognition) are distinct from those used to process information using ambient vision (required for orientation and egomotion). It has even been suggested that processing ambient visual information is pre-attentive (i.e., does not require attentional resources) [13].

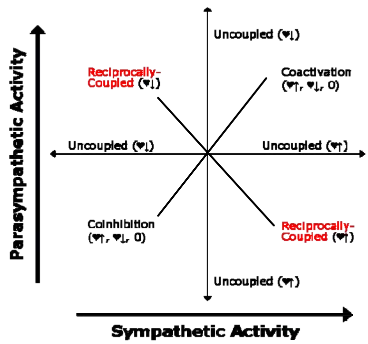


Fig. 1. The eight modes of autonomic control. Traditional coupled modes along the negative diagonal. ♥ = heart rate. Coinhibition and coactivation have multiple responses for heart rate depending on the relative amount of activation or inhibition.

1.3 Deciphering Psychological-Physiological Mappings

A physiological measure is diagnostic to the extent that the measure exhibits a one-to-one psychological-physiological mapping [14]. Previous research looking at heart rate has shown that central processing tasks (e.g., mental arithmetic) elicit reciprocally-coupled SNS activation and PNS withdrawal [15]. Tasks that require primarily perceptual/manual processing tasks (e.g., compensatory tracking) elicit uncoupled PNS withdrawal [16]. In addition, Backs et al. [17] found that combining tasks that require primarily perceptual/manual processing with tasks that require primarily central processing increased PNS withdrawal but did not affect SNS activity. Finally, Backs [17-18] suggests that executive attentional processes (for higher-order strategies during dual-task performance) elicit uncoupled SNS activation.

Recent research has confirmed that these modes of autonomic control are often elicited by facets of the driving task as well. Backs et al. [19] were able to demonstrate that simulated driving through curves of varying degrees (primarily a perceptual/manual processing task) elicited uncoupled PNS withdrawal. In addition, Lenneman et al. [20] were able to show that performing a simulated driving task elicited uncoupled PNS withdrawal, and simulated driving with a verbal working memory task (primarily a perceptual/central processing task) elicited reciprocally-coupled SNS activation and PNS withdrawal. Thus, we believe that the autonomic space for heart rate can provide information about the attentional resources required during driving and performance of a concurrent side task.

1.4 Purpose of the Study

The purpose of the study was to test the diagnosticity of autonomic space for driving research by elucidating psychological-physiological mappings for heart rate. We predict that the *n*-back task will elicit a shortening of PEP and a decrease in RSA because working memory (central processing) resources elicit reciprocally-coupled SNS activation and PNS withdrawal compared to resting baseline. Further, we predict that the magnitude of change in reciprocally-coupled SNS activation and PNS withdrawal will increase as the amount of central processing resources needed to perform the *n*-back task increases from 0-back to 3-back.

Regarding the dual-task conditions, we predict that adding the simulated driving task to the *n*-back task will elicit a change in SNS and PNS activity compared to single-task *n*-back. RSA will decrease (PNS withdrawal) from single-task *n*-back to dual-task *n*-back with driving because of the increase in perceptual/manual processing resources needed to perform simulated driving (i.e., lane keeping). PEP will shorten (SNS activation) from single-task *n*-back to dual-task *n*-back with driving because of the increase in executive processes needed to concurrently perform the two tasks.

2 Method

2.1 Participants

Thirty-two participants (16 female) who were in good health and were not taking any medications that affect the cardiovascular system, were recruited through the Psychology Department subject pool. Participants ranged in age from 18 to 34 (mean = 19.8) years-old. The number of km driven per year for the participants ranged from 3,200 to 51,200 km (mean = 13.920 km).

2.2 Apparatus

The electro- (ECG) and impedance (ICG) cardiograms were obtained from a Minnesota Model 304b (2 participants) or a Mindware Model 2000 cardiograph using Mindware Acq (Mindware Technologies, Inc. ver 2.0) data acquisition system. A desktop DriveSafety driving simulator running HyperDrive (ver. 1.9) was used along with a second computer connected to the simulator to present the *n*-back task.

2.3 Procedure

Participants performed two tasks. One task was simulated driving in which participants had to steer but velocity was controlled by the computer. The driving environment depicted in the simulation was a straight, two-lane road with no traffic ahead (see Fig. 2). Varying levels of wind disturbance were applied during each trial that was simulated as a random amount of force applied perpendicular to the direction of travel from either the left or the right. In the other task participants performed a verbal working memory *n*-back task in which the amount of attentional resources required to perform the task increases as the number of letters to be remembered between the stimulus and target increases. During the 0-back task, the participant

indicated whether the current letter (the stimulus) was the same as the first letter presented at the beginning of the simulation run (the target). During the 3-back task, the participant indicated whether the stimulus matched the letter that was presented three letters previously (the target; see Fig. 2). During both *n*-back tasks, the participants signaled whether the stimulus and target matched or not by pressing different buttons on the steering wheel. The stimuli for the *n*-back task (angular consonants) were presented on road signs placed every 90 m. Thus, when traveling at 72.4 kph (45 mph), the stimuli were presented to the driver every 4.5 s. A total of 53 letters were presented to the participant during each simulation run.

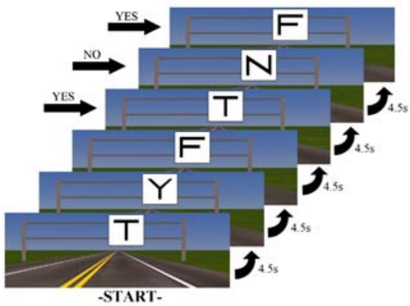


Fig. 2. A 3-back task presented on road signs in the simulated environment. Arrows on the left indicate the correct response for each letter.

At the beginning of the session, participants practiced each task. After practice participants were given a 10-min break while electrodes were applied. Next, an 8-min resting baseline was collected followed by a trial of the driving-only task, which was identical to an *n*-back condition except that no letters were presented on the road signs. Then there were four *n*-back trials (single-task 0- and 3-back, dual-task 0- and 3-back with driving) presented in an order specified by a Latin square, and the order was repeated three times for a total of 12 test trials. During single-task *n*-back, the driving simulation was still presented to the driver; however, steering, braking and acceleration were all computer-controlled. At the end of the test trials participants received a second driving-only trial followed by a second 8-min resting baseline. The present study only used the data from the resting baseline and the 12 test trials to examine the effect of adding simulated driving to *n*-back task performance. The effect of adding the *n*-back task to simulated driving was examined in a previous study using the resting baseline, driving only, and *n*-back dual-tasks [21].

2.4 Data Quantification

ECG and ICG were used to obtain noninvasive indices of SNS and PNS activity. PEP (the SNS index), which is the time between the onset of ventricular depolarization and the onset of left ventricular ejection into the aorta (i.e., Q-B time in ms), was obtained from the first derivative of pulsatile changes in transthoracic impedance (dZ/dt) using Mindware IMP (ver. 2.2, Mindware Technologies, Inc.). RSA (the PNS index) was calculated as the natural logarithm of the variance in the high-frequency heart period

variability frequency band (0.12-0.40 Hz) by applying FFT to the resampled R-R intervals using Mindware HRV (ver. 2.2, Mindware Technologies, Inc.). Heart period was analyzed instead of heart rate because of its superior biometric properties and was calculated as the time in ms between successive R-peaks [22]. A shortening of heart period reflects an increase in heart rate. Respiration rate (in breaths/min) and amplitude (in arbitrary units) were obtained from the dZ/dt data [23]. Physiological data were analyzed in two segments. A mean resting baseline for each measure was calculated across both resting baseline trials, and difference scores for each trial were calculated as the change between the raw score of a measure and the average resting baseline score for that measure. Positive scores indicate an increase in the measure from baseline-to-task (longer heart periods (i.e., slower heart rate), longer PEP, and increased RSA), while negative scores indicate a decrease in the measure from baseline-to-task (shorter heart periods (i.e., faster heart rate), shorter PEP, and suppressed RSA). Performance on the *n*-back task was measured as reaction time in ms from the stimulus presentation to the participant's key press and as the proportion correct in each trial.

3 Results

To test our predictions regarding the effects of single-task *n*-back and dual-task driving and *n*-back, *t*-tests were conducted for heart period, PEP, RSA, and respiration rate and amplitude difference scores. These *t*-tests revealed that both dual-task driving with *n*-back conditions elicited reciprocally-coupled SNS activation and PNS withdrawal and the single-task 3-back approached significance. PEP and RSA decreased from resting baseline to dual-task *n*-back with driving and single-task 3-back. Single-task 0-back did not elicit any physiological change compared to resting baseline.

To test our predictions regarding the effects of adding simulated driving to *n*-back task performance a 2 (attention: single- or dual-task) x 2 (memory load: 0- or 3-back)

Table 1. Means and Standard Errors for Physiological Difference Scores for the Single-task *n*-back and Dual-task Driving and *n*-back. N = 32.

Measure	Single-task 0-back	Single-task 3-back	Dual-task 0-back with driving	Dual-task 3-back with driving
Cardiac				
Heart Period (ms)	-8.38	-47.59***	-19.90**	-55.80***
	(5.58)	(8.81)	(6.10)	(7.64)
PEP (ms)	-1.62	-2.72*	-2.58**	-3.79**
	(0.98)	(1.53)	(1.26)	(1.37)
RSA ($\ln(ms)^2$)	-0.03	-0.34**	-0.25**	-0.48**
	(0.06)	(0.11)	(0.07)	(0.09)
Respiration				
Rate (breaths/min)	0.70	0.14	0.76	-0.09
	(0.60)	(0.58)	(0.48)	(0.58)
Amplitude (arbitrary units)	-0.02	-0.001	-0.003	-0.014
	(0.004)	(0.011)	(0.009)	(0.017)

* = p < .10, ** = p < .05, ***= p < .01.

x 3 (trial) x 2 (segment) repeated-measures ANOVA was conducted for heart period, PEP, RSA, respiration rate, and respiration amplitude difference scores. (There were few trial or segment effects which did not affect the interpretation of the task manipulations and will not be discussed further.) The main effect of attention was significant for heart period, $F(1,31) = 10.38$, $p < .01$, and RSA, $F(1,31) = 14.24$, $p < .01$, and PEP approached significance, $F(1,31) = 4.09$, $p < .053$,. Heart period, PEP, and RSA decreased as a result of adding driving to *n*-back (see Table 2). The main effect of memory load was significant for heart period, $F(1,31) = 41.97$, $p < .001$, and RSA, $F(1,31) = 14.24$, $p < .01$, but not for PEP. Heart period and RSA decreased as memory load increased. The main effect of attention was not significant for respiration rate or respiration amplitude. The main effect of attention was not significant for respiration rate or respiration amplitude, but the main effect of memory approached significance for respiration rate, $F(1,31) = 3.12$, $p < .10$. Thus, the RSA changes with the task manipulations were not merely a function of respiratory change.

The results generally supported our prediction that adding the driving task to the *n*-back task would elicit reciprocally-coupled SNS activation and PNS withdrawal. The results did not support our prediction that the increase in difficulty from 0-back to 3-back would elicit reciprocally-coupled SNS activation and PNS withdrawal. Though the change in PEP was in the right direction, the increase in *n*-back difficulty only elicited significant PNS withdrawal (a decrease in RSA).

The main effect of attention was significant for reaction time, $F(1,31) = 5.51$, $p < .05$, but not for accuracy. Reaction time decreased from single- to dual-task conditions. The main effect of memory was significant for both reaction time, $F(1,31) = 67.69$, $p < .001$, and accuracy, $F(1,31) = 17.97$, $p < .001$. Reaction time increased and accuracy decreased from the 0-back to the 3-back task.

Table 2. Means and Standard Errors for Physiological Difference Scores and Performance for the Factors of Attention and Memory Load. N = 32.

	Attention		Memory Load	
Measure	Single-task	Dual-task	0-back	3-back
Cardiac				
HP (ms)*** $^{+++}$	-27.99	-37.85	-14.14	-51.69
	(6.31)	(6.47)	(5.46)	(7.99)
PEP (ms)*	-2.17	-3.03	-1.95	-3.25
	(1.07)	(1.27)	(0.97)	(1.43)
RSA (ln(ms)^2)*** $^{+++}$	-0.19	-0.36	-0.14	-0.41
	(0.07)	(0.07)	(0.06)	(0.10)
Respiration				
Rate (breaths/min)+	0.31	0.34	0.62	0.03
	(0.52)	(0.50)	(0.49)	(0.59)
Amplitude (arbitrary units)	-0.004	-0.005	-0.002	-0.007
	(0.012)	(0.012)	(0.011)	(0.013)
N-back Performance				
Reaction Time (s)**$^{+++}$	0.74	0.73	0.61	0.86
	(0.03)	(0.03)	(0.03)	(0.04)
Accuracy^{+++}	0.94	0.94	0.98	0.89
	(0.01)	(0.01)	(0.01)	(0.02)

Significant attention effect: * = p < .10, ** = p < .05, ***= p < .01.
Significant memory effect: $^{+}$ = p < .10, $^{++}$ = p < .05, $^{+++}$ = p < .01.

4 Discussion

In general, we were able to successfully demonstrate that cardiac measures are diagnostic to the source of the attentional demands during simulated driving with a verbal working memory side-task. Heart period and RSA were consistent with our predictions, especially at higher levels of difficulty like the single-task 3-back and the dual-tasks. However, although PEP tended to be consistent with our predictions it did not achieve statistical significance in several notable instances. We offer three explanations for why the predicted modes of autonomic control were not elicited.

First, we believe that there may have been little resource competition between simulated driving and the 0-back task. One reason for the lack of resource competition may be because the tasks are data-limited. This contention is supported by the failure of the 0-back task to elicit a significant change in heart period, SNS or PNS activity. Another reason for the lack of resource competition may be because the tasks draw upon different visual channels. For example, the lane keeping task used in our study may only require ambient resources whereas the *n*-back task requires focal resources [24]. Thus, the need executive processes to coordinate task performance would be minimal, making the elicitation of a significant activity of the SNS unlikely.

Second, although executive processes may have been required at the beginning of the experimental session, these may have become more efficient over the course of the experiment [25-26]. In the current study, an increase in efficiency of executive processes may have begun during practice so that by the time of testing minimal executive processing may have been needed. Previous research has shown that SNS activation can wane over time as executive processing becomes more efficient [17].

Third, we cannot discount measurement errors that may have reduced the power of our study, especially for effects upon PEP. PEP change with task manipulations often approaches the single ms level of accuracy associated with this measure, and the B-point can be difficult to reliably identify in some individuals [27].

In summary, our ability to predict the modes of autonomic control provides evidence that cardiac measures can be diagnostic as to the source of the attentional demand during dual-task driving and side-task performance. However, research is needed to examine the sensitivity of the PEP measure of SNS activity with regard to the development of executive processing efficiency and resource competition between focal and ambient visual channels. In addition, research into alternative methods for reducing physiological data (particularly PEP) should be validated in the driving domain. Finally, future research on attentional processing resource competition in simulated driving may want to avoid the use of the 0-back task and use at least a 1-back task to ensure the need for executive processes during dual-task performance.

Acknowledgments. The authors would like to thank Jonathon Shelley for his invaluable programming assistance.

References

1. Wang, J.S., Knipling, R.R., Goodman, M.J.: The Role of Driver Inattention in Crashes: New Statistics From the 1995 Crashworthiness Data System. 40th Annual Proceedings of the Association for the Advancement of Automotive Medicine. Vancouver, British Columbia. pp. 377–392 (1996)
2. Stutts, J.C., Reinfurt, D.W., Staplin, L., Rodgman, E.A.: The Role of Driver Distraction in Traffic Crashes. University of North Carolina Highway Safety research Center, Chapel Hill, NC (2001)
3. Lee, J.D., Strayer, D.L.: Preface to the Special Section on Driver Distraction. Human Factors 46, 583–586 (2004)
4. Kramer, A.: Physiological Metrics of Mental Workload: A Review of Recent Progress. In: Damos, D.L. (ed.) Multiple-task Performance. pp. 279–328. Taylor & Francis, London (1991)
5. O'Donnell, R.D., Eggemeier, F.T.: Workload Assessment Methodology. In: Boff, K.R., Kaufman, L., Thomas, J.P. (eds.) Handbook of Perception and Human Performance, Vol. II, Cognitive Processes and Performance, vol. 42(1), pp. 42–49. Wiley Interscience, New York (1986)
6. Cannon, W.B.: The Wisdom of the Body. Norton, New York (1939)
7. Berntson, G.G., Cacioppo, J.T., Quigley, K.S.: Autonomic Determinism: The Modes of Autonomic Control, the Doctrine of Autonomic Space, and the Laws of Autonomic Constraint. Psychol. Rev. 98, 459–487 (1991)
8. Berntson, G.G., Cacioppo, J.T., Quigley, K.S.: Cardiac Psychophysiology and Autonomic Space in Humans: Empirical Perspectives and Conceptual Implications. Psychol. Bull. 114, 296–322 (1993)
9. Berntson, G.G., Cacioppo, J.T., Quigley, K.S., Fabro, V.T.: Autonomic Space and Psychophysiological Response. Psychophysiology 31, 44–61 (1994)
10. Backs, R.W.: An Autonomic Space Approach to the Psychophysiological Assessment of Mental Workload. In: Hancock, P.A., Desmond, P.A. (eds.) Stress, Workload, and Fatigue, pp. 279–289. Lawrence Erlbaum Associates, Mahwah (2001)
11. Wickens, C.D.: Multiple Resources and Performance Prediction. Theor. Issues Ergonom. Sci. 3, 159–177 (2002)
12. Cacioppo, J.T., Tassinary, L.G.: Inferring Psychological Significance from Physiological Signals. Am. Psychol. 45, 16–28 (1990)
13. Previc, F.H.: The neuropsychology of 3-D space. Psycholog. Bull. 124, 123–164 (1998)
14. Cacioppo, J.T., Tassinary, L.G.: Inferring Psychological Significance From Physiological Signals. Am. Psychol. 45, 16–28 (1990)
15. Berntson, G.G., Cacioppo, J.T., Fieldstone, A.: Illusions, Arithmetic, and Bidirectional Modulation of Vagal Control of the Heart. Biol. Psychol. 44, 1–17 (1996)
16. Lenneman, J.K., Backs, R.W.: The Validity of Factor Analytically Derived Autonomic Components for Mental Workload Assessment. In: Backs, R.W., Boucsein, W. (eds.) Engineering Psychophysiology: Issues and Applications, pp. 161–175. Lawrence Erlbaum Associates, Mahwah, NJ (2000)
17. Backs, R.W., Rhody, J., Barnard, J.: Cardiac Control During Dual-Task Performance of Visual or Auditory Monitoring With Visual-Manual Tracking. Psychologia 48, 66–83 (2005)
18. Backs, R.W.: A Comparison of Factor Analytic Methods of Obtaining Cardiovascular Autonomic Components for the Assessment of Mental Workload. Ergonomics 41, 733–745 (1998)

19. Backs, R.W., Lenneman, J.K., Wetzel, J.M., Green, P.: Cardiac Measures of Driver Workload During Simulated Driving With and Without Visual Occlusion. Human Factors 45, 525–538 (2003)
20. Lenneman, J.K., Shelley, J.R., Backs, R.W.: Deciphering Psychological-Physiological Mappings During a Simulated Driving Task. In: Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design. Rockland, ME, pp. 493–498 (2005)
21. Lenneman, J.K., Tuttle, S.J., Oliver, M.L., Backs, R.W.: Cardiac Autonomic Control During Dual-Task Performance of Simulated Driving and a Verbal Working Memory Task. In: Proceedings IEA2006 Congress, Elsevier, Ltd, Oxford, U.K (2006)
22. Berntson, G.G., Cacioppo, J.T., Quigley, K.S.: The Metrics of Cardiac Chronotropism: Biometric Perspectives. Psychophysiology 32, 162–171 (1995)
23. Ernst, J.M., Litvack, D.A., Lozano, D.L., Cacioppo, J.T., Berntson, G.G.: Impedance Pneumography: Noise as Signal in Impedance Cardiography. Psychophysiology 36, 333–338 (1999)
24. Horrey, W.J., Wickens, C.D., Consalus, K.P.: Modeling Driver's Visual Attention While Interacting With In-Vehicle Technologies. J. Exp. Psychol. Appl. 12, 67–78 (2006)
25. Fisk, A.D., Schneider, W.: Memory as a Function of Attention, Level of Processing, and Automatization. J. Exp. Psychol. Learning, Memory, and Cognition. 10, 181–197 (1984)
26. Schneider, W., Fisk, A.D.: Degree of Consistent Training: Improvements in Search Performance and Automatic Process Development. Percept. Psychophys 31, 160–168 (1982)
27. Lozano, D.L., Norman, G., Knox, D., Wood, B.L., Miller, B.D., Emery, C.F., Berntson, G.G.: Where to B in dZ/dt. Psychophysiology 44, 113–119 (2007)

EEG Activities of Dynamic Stimulation in VR Driving Motion Simulator

Chin-Teng Lin[1,3], Li-Wei Ko[1,3], Yin-Hung Lin[1], Tzyy-Ping Jung[1,2], Sheng-Fu Liang[1,4], and Li-Sor Hsiao[1,3]

[1] Brain Research Center, National Chiao-Tung University (NCTU), Hsinchu 300, Taiwan, R.O.C

[2] Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California, San Diego, USA

[3] Department of Electrical and Control Engineering, National Chiao-Tung University (NCTU), Hsinchu 300, Taiwan, R.O.C

[4] Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan 701, Taiwan, R.O.C

Abstract. The purpose of this study is to investigate Electroencephalography dynamics in response to kinesthetic stimuli during driving. We used a Virtual Reality driving simulator consisted of a hydraulic hexapod motion platform to create practical driving events. We compared the EEG dynamics in response to kinesthetic stimulus while the platform was in motion, to that while the platform was stationary. The scalp-recorded EEG channel signals were first separated into independent brain sources using Independent Component Analysis (ICA), and then studied with time-frequency analysis. Our results showed that independent brain processes near the somatomotor cortex exhibited alpha power decreases across sessions and subjects. Negative potentials phase-locked to the onsets of deviation events under motion conditions were observed in a central midline component. The results allow us to better understand different brain networks involved in driving, and provide a foundation for studying event-related EEG activities in the presence of kinesthetic stimuli.

Keywords: Kinesthetic Stimulus, EEG, ICA, Component Clustering, ERSP, ERP, Mu Rhythm, EMG.

1 Introduction

Kinesthetic perception -- the sensory apparatus that detects motion -- is one of the most important sensations to human beings, yet we usually overlook the contributions of the vestibular system to our lives. However, we would not have a complete sensation of the world without the perception of motion. We cannot even stand still or walk in a straight line without the vestibular system functioning properly. The vestibular system thus plays an important role in our lives.

Researchers have tried to measure evoked potentials of vestibular origin for 30 years. Elidan et al. [1] reported the ERP response to high speed and transient vertical Z axis rotation. Subjects were rotated at the speed of 10,000°/sec2 for 2 ms. The

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 551–560, 2007.

reported negativity peaked at about 15 ms after the onsets of rotation from signals measured at a forehead mastoid electrode. Baudonniere et al. [2] reported a biphasic negative wave, that is most prominent at central midline electrode (Cz) in subjects who received short (30 ms) linear displacements without co-stimulation of the semicircular canals.

Probst et al. [3-5] and Loose et al. [6] observed bell-shaped negativity at central midline channels following roll up and down motion along the X axis. The Vestibular Evoked Potential (VESTEP) evoked by stimulating otolithic and semicircular canals with different orientations of rotations or directions of movements was investigated in depth.

The experimental variables in these studies were well controlled. This might be desirable from the perspective of scientific research, but is less practical because we rarely experience vestibular stimulation without visual co-stimulation or watch pixels rotating or moving in the real world. We actually live in a visual-vestibular co-stimulation world and the visual cue is always a meaningful and continuous scene -- in driving, for instance.

The driving perception includes the co-stimulation of visual cue, vestibular stimulation, muscle reaction, and skin pressure. This is indeed a complicated mechanism to understand.

Using a realistic simulator to conduct driving experiments is widely used in driving-related research [7]. Regarding the necessity of motion during driving, the literature shows that the absence of motion information increases reaction times to external movement perturbations [8], and decreases safety margins in the control of lateral acceleration in curve driving [9]. In real driving, improper signals from disordered vestibular organs were reported to contribute to inappropriate steering adjustment [10]. Groen et al. [11] also showed that the presence of vestibular information in driving simulators was important in the perception of illusory self-tilt and illusory self-motion. These studies emphasized the importance of motion perception during driving to the assessment of driving performance and behavior.

The electroencephalogram (EEG) is a popular method for evaluating human cognition. Compared to functional Magnetic Resonance Imaging (fMRI), EEG is much less expensive and more portable, thus it is applicable in our daily lives, especially on the move.

In recent years, researchers have designed the Virtual Reality (VR) scenes to provide appropriate environments for assessing brain activity during driving [12-14]. Lin et al. [12,13] introduced the “dynamic” VR environment in conjunction with physiological and behavioral response recordings to offer more assessment options than were available in traditional neuropsychological studies. However, the EEG correlates of kinesthetic stimulations induced by the motion platform in the dynamic VR scene have not been fully assessed or appreciated.

The purpose of this study is to investigate EEG dynamics in response to kinesthetic stimuli using a dynamic VR environment. To this end, we constructed an interactive driving environment that integrated a surrounding scene and a real vehicle mounted on a hydraulic hexapod motion platform. This dynamic VR environment mimicked visual-vestibular co-stimulation during driving. Using simple driving behaviors, we studied brain responses of kinesthetic inputs by comparing subjects’ EEG differences in motion and motionless conditions of the dynamic platform.

2 Material and Methods

We developed a VR-based 3D high-fidelity interactive highway scene. The synchronized scenes were projected from seven projectors to constitute a surrounding vision. At the center of the projected scenes, a real vehicle mounted on the motion platform to provide motion sensations. The vestibular cues were delivered by a Stewart Platform [15]. The platform generated accelerations in vertical, lateral, and longitudinal directions of a vehicle as well as pitch, roll, and yaw angular accelerations. This technique has been used widely in driving simulation studies [16].

We designed three driving events: stop, go, and deviation. Figure 1 shows the time course of a typical Stop-Go event. Subjects did not need to do anything in the events. Moreover, the vehicle was randomly drifted away from the cruising position, and the subjects were instructed to steer the vehicle back to the center of the cruising lane as quickly as possible (Figure 2).

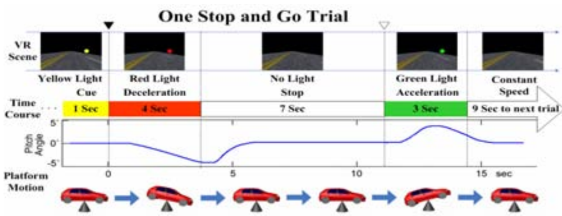


Fig. 1. Illustration of the design for Stop-Go events in driving

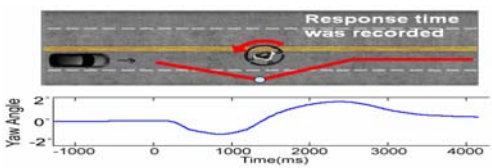


Fig. 2. The vehicle was randomly drifted away from the cruising position, which was defined as a deviation event, and the subjects were instructed to steer the vehicle back to the center of the cruising lane as quickly as possible

Ten healthy subjects participated in this research (aged between 20 and 28). Subjects were instructed to keep the car at the center of the inside lane by controlling the steering wheel, and to perform the driving task consciously. Each subject completed four 25-minute sessions in each driving experiment. The entire driving experiment lasted about 2 hours. Subjects performed at least 2 driving experiments on different days for testing the cross-session consistency.

The physiological data acquisition uses 33 unipolar sintered Ag/AgCl EEG/EOG electrodes and 2 bipolar ECG electrodes placed on the chest. All the EEG/EOG electrodes were placed based on a modified International 10-20 system and refer to the right ear lobe. The contact impedance between EEG electrodes and scalp was calibrated to be less than 5kΩ. We used the Scan NuAmps Express system

(Compumedics Ltd., VIC, Australia) to simultaneously record the EEG/EOG/ECG data and the deviation between the center of the vehicle and the center of the cruising lane triggered by the VR program. The EEG data were sampled at 500 Hz with a 16-bit quantization level.

The continuous EEG signals were first extracted into epochs whose lengths were designed to cover the whole platform dynamics in single driving events. We then applied Independent Component Analysis to concatenated epochs to decompose them into temporally statistical component activations.

ICA methods have been extensively applied to the blind source separation problem since the 1990s [17-21]. Subsequent technical reports [22-28] demonstrated that ICA was a suitable solution to the problem of EEG source segregation, identification, and localization.

To study the cross-subject component stability of ICA decomposition, components from multiple sessions and subjects were clustered based on their spatial distributions and EEG characteristics [24], [29].

Component Clustering grouped massive components from multiple sessions and subjects into several significant clusters and identified at least 9 clusters of components having similar power spectra and scalp projections. These component clusters also showed functionally distinct activity patterns.

Single-trial event-related potential (ERP) data are usually averaged prior to analysis to increase their signal/noise relative to non-time and -phase locked electroencephalographic (EEG) activity and non-neural artifacts.

Event-Related Potential (ERP) images directly visualized single event-related EEG trials and their contributions to the averaged ERP [24]. An ERP image also makes visible relationships between subject behavior and amplitudes/latencies of individual event-related responses.

Event-Related Spectral Perturbation (ERSP) plots the grand mean time course of changes from pre-stimulus baseline in log spectral power of a scalp-recorded EEG or ICA component activation time-locked to stimulus presentation or subject responses across frequencies. Through ERSP, we are able to observe time-locked but not necessarily phase-locked activities [30].

3 Results

3.1 Mu Component Activations

Figure 3 shows the ERSP of a component that exhibited differential brain responses between motion and motionless conditions. The component scalp map exhibited the defining features of mu rhythms -- distinct spectral peaks near 10 Hz and 22 Hz. The upper ERSP panels show the ERSP following stop events, while the lower two show go-event ERSP. Images on the left are brain responses under “motion” conditions and those on the right are under “motionless” conditions. The curves below the images are the time courses of the platform motion (pitching, rotate by Y axis). Mu power was strongly blocked (reduced) around the peak of platform motion in Motion-Stop and Motion-Go events. In contrast, no mu blocking occurred following either stop or go events in the motionless condition (Fig. 3, right panels). Thus the mu blocking appears to be induced by the kinesthetic inputs in stop and go events.

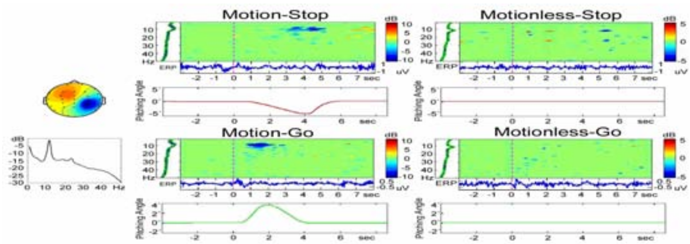


Fig. 3. A right mu component shows mu characteristic 10 Hz and 22 Hz peaks in the activity spectrum (lower left). The component mean ERSP shows mean event-related changes in (log) spectral power across data trials time-locked to the kinesthetic stimulus onsets (dashed line). Following the motion platform movement, this activity is blocked. The activity was unchanged from the baseline spectra if the motion platform was not in action (right).

Mu blocking was also observed following deviation events. Figure 4 shows the ERSP of a right mu component following deviation events. The upper and lower panels show ERSPs of the component following "deviate-to-left" and "deviate-to-right" events, respectively. The curves below show the motion recordings of the platform. Notice that the motion platform tilted along different directions in Stop-Go and deviation events (cf. Fig. 3). In deviation events, the platform rotated slightly along the vertical Z axis.

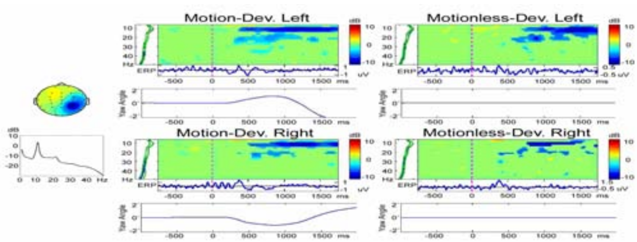


Fig. 4. The mean ERSP of the mu component follow deviation events

When deviation occurred, the subjects were instructed to maneuver the car back to the cruising position by steering the wheel. It is expected that mu activity would be blocked due to the hand movement in both motion and motionless conditions. However, the latency of mu blocking in the motion condition was significantly shorter than that in the motionless condition.

3.2 Central Midline (CM) Component Activations

Figure 5 shows the scalp map and dynamic properties of an independent component from the same subject in Motion-Deviation and Motionless-Deviation conditions. The scalp map of the CM component (Fig. 5 upper left) resembles scalp maps of the "P3a" or "P3novel" ERP peaks [31-32]. In two-dimensional "ERP image" plots of single trials from the subject, potential fluctuations are shown as color-coded horizontal lines, here normalized by component activation baseline variability then sorted by response time (RT). The ERP images clearly show that the early kinesthetic response,

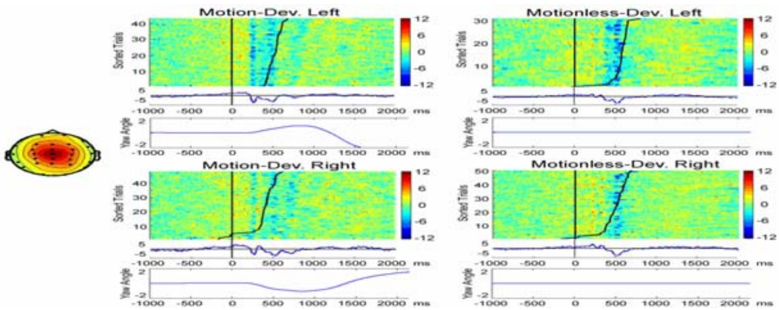


Fig. 5. Single-trial Event-Related Potentials (ERPs) of the central midline (CM) component

peaked at ~250 ms, was time-locked to deviation onset. However, this sharp negativity was missing in the motionless condition (Fig. 5 right panels).

3.3 Component Stability

There were 29 components from 10 subjects contributing to a large cluster of left mu components. We found that the 11 Hz activity was blocked following kinesthetic stimuli in the motion condition. Hence, we strongly suggested that these represented mu activity [33]. Scalp maps of individual left mu components in this cluster strongly resembled the cluster mean map.

Figure 6a shows the component cluster mean ERSP of the component activations following Stop-Go events under the motion and motionless conditions. The ERSP images of motion sessions exhibited a strong mu blocking and alpha rebound, which were completely missing from the ERSP images under motionless conditions, consistent with the results in a typical subject shown in Figure 3.

Similarly, Figure 6b shows averaged ERSP images following deviation events under motion and motionless conditions. Although the ERSP images in all four conditions exhibited similar mu blocking induced by the steering actions, the latencies of mu blocking differed considerably.

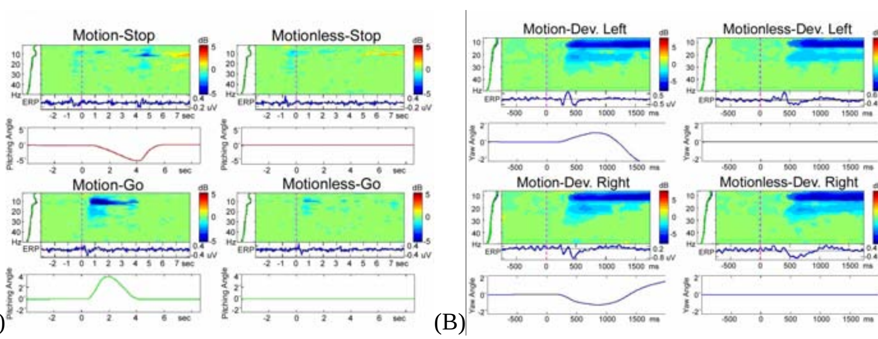


Fig. 6. (A) The group-averaged ERSP shows the component activations following Stop-Go events under the motion (left panels) and motionless (right panels) conditions. (B) The group-averaged ERSP images following deviation events under motion and motionless conditions.

4 Discussions

In this study, we recorded and analyzed unaveraged single-trial EEG data in 31 driving experiments from 10 volunteer drivers under two different driving conditions -- motion and motionless. The hexapod motion platform that simulated driving events allowed us to study neural correlates of kinesthetic stimuli. We performed ICA to separate the EEG contributions of distinct brain processes to explore their individual and joint event-related dynamics following Stop-Go and deviation events through ERP differences and time-frequency analysis (ERSP). Alpha power of the mu component cluster was strongly blocked (~-5dB) around the peak of platform movement in Motion-Stop and Motion-Go events. A sharp negative was found in the central midline component cluster only in Motion-Deviation events. We believe that these two features were induced by kinesthetic stimuli.

4.1 Mu Components

Mu rhythm is an EEG rhythm recorded usually from the motor cortex of the dominant hemisphere. It is a variant of normality, and it can be suppressed by a simple motor activity, or passively moved [6, 34, 35].

Deviation events involved subject responses to steer the vehicle back to the cruising position. Thus it is expected that mu power would be blocked following deviation events. Our results showed unexpected strong mu blocking in response to Motion-Stop and Motion-Go events in which no action was involved, suggesting kinesthetic stimuli could also induce mu blocking.

Following deviation events, mu power was strongly blocked in both motion and motionless conditions. We found the latency of mu blocking in Motion-Deviation events would lead that in Motionless-Deviation events by a comparable length. Mu blocking thus appeared associated with kinesthetic stimuli delivered to the drivers. In short, long-lasting mu blocking following deviation events began with the EEG brain dynamics induced by kinesthetic stimuli, followed by marked mu power decrease associated with subject motor actions.

4.2 Central Midline Components

The central midline component cluster exhibits a sharp negativity in averaged ERP following Motion-Deviations, but the negativity is missing from the ERP following Motionless-Deviations.

The sharp negativity in the ERP of the central midline component cluster is also consistent with previous VESTEP studies of Elidan et al. [36-38]. They showed a negative potential near Cz or forehead, induced by external kinesthetic stimulus. However, they did not report any mu blocking in response to the kinesthetic stimuli. To the best of our knowledge, this finding had never been reported in the past.

4.3 Alpha Activity and Drowsiness

Traditionally, the EEG alpha band was used as an indicator of drowsiness estimation during driving [12-14]. The alpha power had been reported to index the level of

drowsiness in attention-sustained experiments in a laboratory setting. In this study, our results showed that alpha-band activity varies during driving, especially when the vehicle was moving and delivered kinesthetic stimuli to the drivers and passengers, which might confound the fatigue-related alpha power changes in driving. Thus, more care must be taken to examine the validity of using alpha power to index drowsiness level in real driving.

References

1. Elidan, J., Leibner, E., Freeman, S., Sela, M., Nitzan, M., Sohmer, H.: Short and middle latency vestibular evoked responses to acceleration in man. Electroencephalogr Clin. Neurophysiol 80, 140–145 (1991)
2. Baudonniere, P.M., Belkhenchir, S., Lepecq, J.C., Mertz S.: Otolithvestibular-evoked potentials in humans. Intensity, direction of acceleration (Z+, Z-), and BESA modeling of generators. Ann NY Acad Sci. Vol. 871, pp. 384–386 (1999)
3. Probst, T., Dabrowski, H., Liebler, G., Wist, E.: MARDER – Multi-Axes Rotation Device for Experimental Research: a new concept for investigations of the vestibular, oculomotor, and visual systems of humans in three-dimensional space. J. Neuroscience Methods 49, 49–61 (1993)
4. Probst, T., Bablok, E., Dabrowski, H., Dombrowski, J.H., Loose, R., Wist, E.R.: Position and velocity responses from the otoliths and the canals: results from ESA's parabolic flights. Aviat Space Environ Med. 67, 633–639 (1996)
5. Probst, T., Ayan, T., Loose, R., Skrandies, W.: Electrophysiological evidence for direction-specific rotary evoked potentials in human subjects – a topographical study. Neurosci Lett. 239, 97–100 (1997)
6. Loose, R., Probst, T., Tucha, O., Bablok, E., Aschenbrenner, S., Lange, K.W.: Vestibular evoked potentials from the vertical semicircular canals in humans evoked by roll-axis rotation in microgravity and under 1-G. Behavioural Brain Research 134, 131–137 (2002)
7. Kemeny, A., Panerai, F.: Evaluating perception in driving simulation experiments. TRENDS in Cognitive Sciences 7, 31–37 (2003)
8. Wierville, W.W., Casali, J.G., Repa, B.S.: Driver steering reaction time to abrupt-onset crosswind, as measured in a moving-base driving simulator. Hum. Factors 25(1), 103–116 (1983)
9. Reymond, G., Kemeny, A., Droulez, J., Berthoz, A.: Role of lateral acceleration in curve driving: driver model and experiments on a real vehicle and a driving simulator. Hum. Factors. 43, 483–495 (2001)
10. Page, N.G., Gresty, M.A.: Motorist's vestibular disorientation syndrome. J. Neurol. Neurosurg. Psychiatry 48, 729–735 (1985)
11. Groen, E.L., Howard, I.P., Cheung, B.: Influence of body roll on visually induced sensation of self-tilt and rotation. Perception 28, 287–297 (1999)
12. Lin, C.T., Wu, R.C., Jung, T.P., Liang, S.F., Huang, T.Y.: Estimating Driving Performance Based on EEG Spectrum Analysis. EURASIP Journal on Applied Signal Processing 19, 3165–3174 (2005)
13. Lin, C.T., Wu, R.C., Liang, S.F., Chao, W.H., Chen, Y.J., Jung, T.P.: EEG-based drowsiness estimation for safety driving using independent component analysis. IEEE Transactions on Circuits and Systems I. 52, 2726–2738 (2005)

14. Eoh, H.-J., Chung, M.-K., Kim, S.-H.: Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. International Journal of Industrial Ergonomics 35, 307–320 (2005)
15. Stewart, D.: A platform with six degrees of freedom. Proc. Instn Mech. Engr. 180, 371–386 (1965)
16. Reymond, G., Kemeny, A.: Motion cueing in the Renault driving simulator. Veh. Syst. Dynam. 34, 249–259 (2000)
17. Jutten, C., Herault, C.: Blind separation of sources I. An adaptive algorithm based on neuromimetic architecture. Signal Process 24, 1–10 (1991)
18. Comon, P.: Independent component analysis – A new concept? Signal Processing 36, 287–314 (1994)
19. Bell, A.J., Sejnowski, T.J.: An information-maximization approach to blind separation and blind deconvolution. Neural Computation 7, 1129–1159 (1995)
20. Girolami, M.: An alternative perspective on adaptive independent component analysis. Neural Computation 10, 2103–2114 (1998)
21. Lee, T.W., Girolami, M., Sejnowski, T.J.: Independent component analysis using an extended infomax algorithm for mixed sub- and super-Gaussian sources. Neural Computation 11, 606–633 (1999)
22. Jung, T.P., Humphries, C., Lee, T.W., Makeig, S., McKeown, M.J., Iragui, V., Sejnowski, T.J.: Extended ICA removes artifacts from electroencephalographic recordings. Advances in Neural Information Processing Systems 10, 894–900 (1998)
23. Jung, T.P., Makeig, S., Humphries, C., Lee, T.W., McKeown, M.J., Iragui, V., Sejnowski, T.J.: Removing electroencephalographic artifacts by blind source separation. Psychophysiology 37, 163–178 (2000)
24. Jung, T.P., Makeig, S., Westerfield, W., Townsend, J., Courchesne, E., Sejnowski, T.J.: Analysis and visualization of single-trial event-related potentials. Human Brain Mapping 14, 166–185 (2001)
25. Yamazaki, A., Tajima, T., Matsuoka, K.: Convolutive independent component analysis of EEG data. Annual Conference on SICE 2, 1227–1231 (2003)
26. Meyer-Base, A., Auer, D., Wismueller, A.: Topographic independent component analysis for fMRI signal detection. In: Proceedings of the International Joint Conference on Neural Networks. Vol. 1, pp. 601–605 (2003)
27. Naganawa, M., Kimura, Y., Ishii, K., Oda, K., Ishiwata, K., Matani, A.: Extraction of a plasma time-activity curve from dynamic brain pet images based on independent component analysis. IEEE Transactions on Biomedical Engineering 52, 201–210 (2005)
28. Liao, R., Krolik, J.L., McKeown, M.J.: An information-theoretic criterion for intrasubject alignment of FMRI time series: motion corrected independent component analysis. IEEE Transactions on Medical Imaging 24, 29–44 (2005)
29. Makeig, S., Delorme, A., Westerfield, M., Jung, T.-P., Townsend, J., Courchense, E., Sejnowski, T.: Electroencephalographic brain dynamics following visual targets requiring manual responses. PLoS Biology 2(6), 747–762 (2004)
30. Makeig, S.: Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. Electroencephalography and Clinical Neurophysiology 86, 283–293 (1993)
31. Courchesne, E., Hillyard, S.A., Galambos, R.: Stimulus novelty, task relevance and the visual evoked potential in man. Electroencephalogr Clin. Neurophysiol 39, 131–143 (1975)
32. Polich, J., Comerchero, M.: P3a from visual stimuli: typicality, task, and topography. Brain Topogr 15, 141–152 (2003)

33. Kuhlman, N.W.: Functional topography of the human mu rhythm. Electroencephalography and Clinical Neurophysiology 44, 83–93 (1978)
34. Thilo, K.V., Kleinschmidt, A., Gresty, M.A.: Perception of self-motion from peripheral optokinetic stimulation suppresses visual evoked responses to central stimuli. Journal of Neurophysiology 90, 723–730 (2003)
35. Parker, D.E., Phillips, J.O.: Self-motion perception: assessment by real-time computer-generated animations. Applied Ergonomics 32, 31–38 (2001)
36. Elidan, J., Sohmer, H., Nizan, M.: Recording of short latency vestibular evoked potentials to acceleration in rats by means of skin electrodes. Electroencephalogr Clin. Neurophysiol 53, 501–505 (1982)
37. Elidan, J., Sohmer, H., Lev, S., Gay, I.: Short latency vestibular evoked response to acceleration stimuli recorded by skin electrodes. Annual of Otolrhinollaryngol 93, 257–261 (1984)
38. Elidan, J., Langhofer, L., Honrubia, V.: Recording of short-latency vestibular evoked potentials induced by acceleration impulses in experimental animals: current status of the method and its applications. Electroencephalogr Clin. Neurophysiol 68, 58–69 (1987)

Development of a Wireless Embedded Brain - Computer Interface and Its Application on Drowsiness Detection and Warning

Chin-Teng Lin[1,2], Hung-Yi Hsieh[1], Sheng-Fu Liang[1,3], Yu-Chieh Chen[1,2], and Li-Wei Ko[1,2]

[1] Brain Research Center, University System of Taiwan, Hsinchu, Taiwan
[2] Department of Electrical and Control Engineering, National Chiao-Tung University, Hsinchu, Taiwan
[3] Department of Computer Science and Information Engineering, National Cheng-Kung University, Tainan, Taiwan

Abstract. The existing bio-signal monitoring systems are mostly designed for signal recording without the capability of automatic analysis so that their applications are limited. The goal of this paper is to develop a real-time wireless embedded electroencephalogram (EEG) monitoring system that includes multi-channel physiological acquisition, wireless transmission, and an embedded system. The wireless transmission can overcome the inconvenience of wire routing and the embedded multi-task scheduling for the dual-core processing system is developed to realize the real-time processing. The whole system has been applied to detect the driver's drowsiness for demonstration since drowsiness is considered as a serious cause of many traffic accidents. The electroencephalogram (EEG) features changes from wakefulness to drowsiness are extracted to detect the driver's drowsiness and an on-line warning feedback module is applied to avoid disasters caused by fatigue.

Keywords: Brain-Computer Interfaces (BCIs), electroencephalogram (EEG), embedded systems, real-time, wireless.

1 Introduction

There are verities of researches that focused on bio-signals monitoring system. However, most of the existing physiological signal monitor systems can only record the signals without the capability of automatic analysis. There's a tendency to make the developed BCI system more convenient to use. Traditionally, the variations of brain waveforms are measured and analyzed by personal computers (PCs). The complexity of the EEG data requires excellent calculating capabilities for the analyzing procedures. However, the inconvenience of using PCs can limit the applications of designed BCI systems. We need to develop wearable and inexpensive BCI systems—small devices with long battery life that can be carried indoors or outdoors [1]. With the development of embedded system and signal processing technique, there is a tendency to apply the techniques to BCI systems.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 561–567, 2007.

The goal of this paper is to develop a real-time wireless embedded electro-encephalogram monitoring system that includes multi-channel physiological acquisition, wireless transmission, and an embedded system. The proposed system employs signal acquisition and amplification units to collect EEG signals and wireless modules to transmit data to a signal processing embedded system. A multi-task scheduling mechanism was also implemented to increase the signal acquisition and analysis performance. The wireless transmission can overcome the inconvenience of wire routing and an embedded multi-task scheduling for the dual-core processing system is developed to realize the real-time processing. Finally, a real-time drowsiness detection method combined with an on-line warning feedback is implemented in the developed system for demonstration.

This paper is organized as follows. The system architecture of the BCI system is introduced in Section 2, including a multi-channel physiological acquisition and amplification unit, a wireless transmission unit, a dual-core signal processing unit, a remote real-signal display and monitoring unit, and a warning device. The application in driver's drowsiness detection is given in Section 3 and the results are shown in Section 4. The conclusions are summarized in Section 5.

2 System Architecture

The block diagram of the developed EEG-based BCI system is showed in Fig.1, which includes five units: (1) signal acquisition and amplification unit, (2) wireless data transmission unit, (3) signal processing unit, (4) remote system for data storage and real-time display, and (5) warning device.

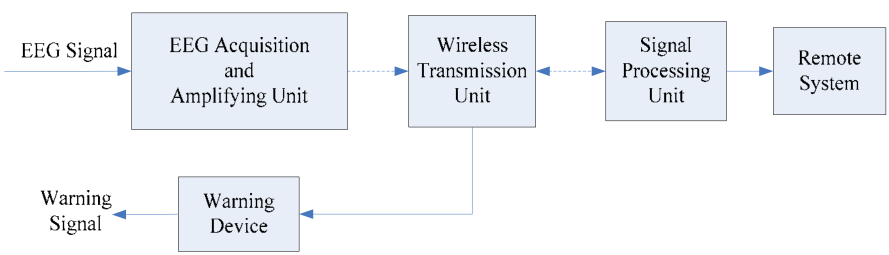


Fig. 1. The block diagram of the proposed BCI system

The data flow of the system can be divided into two streams: (1) The EEG signal was first acquired by signal acquisition and amplification unit, and then transmitted from wireless data transmitter to wireless data receiver. The EEG signals are then processed by the data processing unit and transmitted to the remote system for data storage and real-time display. (2) After the process of EEG data, the system will transmit the result to remote system by TCP/IP and trigger the warning device when the drowsiness condition occurs.

The signal acquisition and amplification unit is applied to measure the EEG signal and to filter the artifacts. The structure is shown in Fig. 2(a), the EEG amplifying circuit consists of a pre-amplifier with the gain of 100, an isolated amplifier to protect

the subject, a band-pass filter to reserve 1-100Hz which was composed of a low-pass filter and a high-pass filter, a differential amplifier which had the gain of ten or fifty (which can be chosen by a switch), and a 60Hz notch filter to eliminate the effect of the line noise.

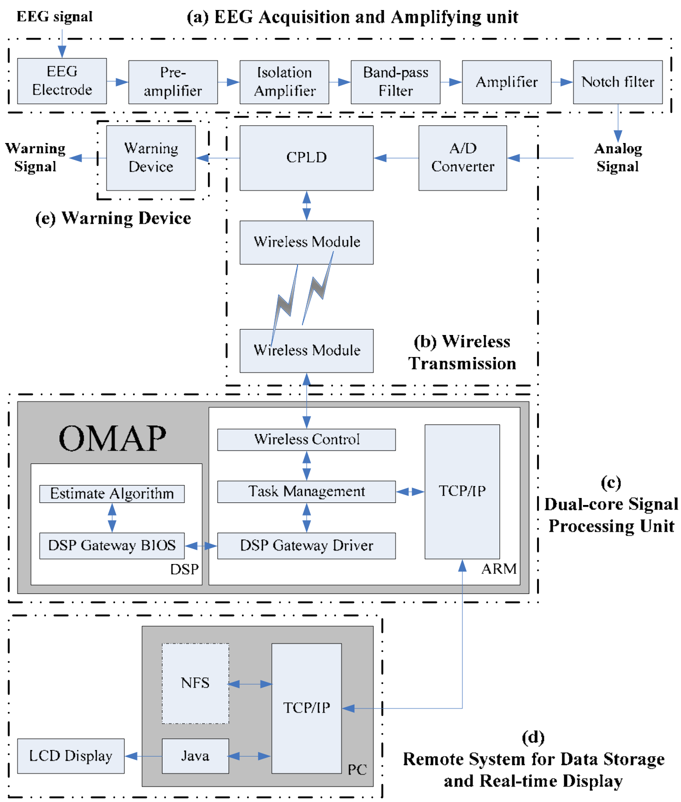


Fig. 2. The detail architecture of the BCI system

Fig. 2(b) shows the wireless data transmission unit which includes an A/D converter, a Complex Programmable Logic Device (CPLD) and wireless modules. The acquired signal is first converted from analog to digital, and then transmitted through the wireless modules. The ALTERA (San Jose, California, USA) FLEX10K EPF10K10TC144-3 CPLD is employed to control the A/D converter and encode the data for the wireless modules. Two different devices can be selected as the wireless module according to transmission distance in the designed BCI system. Bluetooth modules are a frequently choice in medical science with short transmission distance. In our previous study [2], we used Bluetooth in the proposed embedded system. Since the transmission distance was limited with Bluetooth, we currently combined the radio frequency (RF) module (RF3100/3105) in the system for longer operational distance. Because the interface of RF3100/3105 and Bluetooth were identical so we

could simultaneously load the codes of each of these two modules into CPLD and switch between them.

The dual-core processing unit is the main part of the wireless BCI. The operating core is Texas Instruments (TI) Open Multimedia Architecture Platform (OMAP) 1510, which is composed of an ARM925 processor and a TMS320C55x DSP processor. The EEG signals were mostly processed in PCs because of its complexity. The DSP processor is helpful for EEG data processing with a large number of mathematical calculations. In this study, the DSP core was used to process EEG data and the ARM925 was used to communicate with other devices, such as wireless transmission modules and TCP/IP network. The DSP Gateway is used as the cooperation structure for the communication between the two cores since these two cores have different function which is shown in Fig. 2(c). DSP Gateway can be considered as a software layer that connects the ARM and DSP processors. DSP gateway makes ARM core possible to use resource of DSP core by Application Program Interface (API), and works like a small real-time kernel which manages the source and data flow in the DSP core. With this mechanism, the DSP processor is started only when the system needs to process the EEG data. The Linux operating system (OS) is built to manage the resource of ARM core [3]. The structure of the dual-core processing unit is shown in Fig. 2(c), the functions of ARM core can be divided into three parts: (1) wireless receiver control, (2) TCP/IP control, and (3) DSP Gateway driver. The ARM core was selected for these tasks for the reason of its excellent interface control ability. An embedded multi-task scheduling system is used to manage these tasks. The multi-task scheduling mechanism was developed to ensure the accurate sampling rate for EEG signal acquisition and data process and analysis in real time [4]. With the architecture, the ARM core will not only hold and wait but also keep transmitting data from the wireless module to the display unit when the DSP core is processing EEG signals.

The structure of the remote system is shown in Fig. 2 (d). The remote system has two important functions, including data storage and real-time EEG signal display. The data size of continuous EEG recordings is beyond the storage capacity of the embedded system. Thus, we have implemented a network file system to store EEG signals remotely. Additionally, we built a graphic user interface (GUI) to show in real-time the bio-medical signals. The connection between the remote system and OMAP we used in BCI system is TCP/IP protocol.

3 Real-Time Driver's Drowsiness Detection and Warning

The proposed BCI system is designed to process bio-signals in real-time. And the system is then applied to estimate driver's drowsiness via automatic EEG analysis for demonstration. In order to implement that, we developed a real-time EEG data analysis algorithm and built an experiment environment to test and verify the algorithm.

3.1 Experiment Environment and Experimental Design

A virtual-reality (VR) based highway-driving environment that was developed in our previous studies [5][6] to investigate the changes of drivers' cognitive states during a long-term driving. The VR driving environment includes 3D surround scenes projected by seven projectors and a real car mounted on a 6-degree-of-freedom Stewart platform. During the driving experiments, all scenes are moving according to the displacement of the car and the subject's steering-wheel handling. The driving speed is fixed as 100 km/hr and the car is randomly and automatically drifted away from the center of the cruising lane to mimic the consequences of a non-ideal road surface. We asked the subject to keep the car on the third cruising lane (from left to right). While the subject is alert, his/her response time will be short and deviation of the car will be small; otherwise, the subject's response time will be slow and the car's deviation can be large. In this driving experiment, the VR-based freeway scene provides only one car driving on the road without any other event stimuli to simulate a monotonous and unexciting task that it tends to make drivers fall asleep. The 32-channel electrode cap was used in our study, in which the electrodes are placed according to traditional 10-20 placement system.

(a) (b)

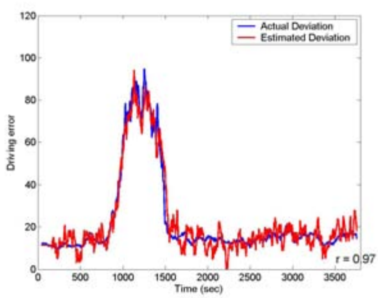
Fig. 3. The virtual-reality scene and the environment

3.2 EEG Signal Analyzing System

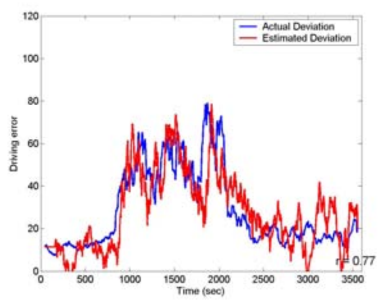
The acquired EEG signals are first down-sampled to 64 Hz to reduce the calculation loading to the system and a Hanning window is then applied to the signals for reducing distortion. The EEG power spectrum transformed by using Short-Time Fourier Transform and then applied to a moving average filter to eliminate noises. We use principal component analysis (PCA) on the normalized EEG power spectrum to reduce data dimension. The features extracted by PCA are finally fed into a linear regression model to estimate the driving performance of the driver. There is no PCA model in our previous studies [5][6], because the algorithm without PCA model can estimate very well in our previous studies. However, in the current setting, changing electrodes position makes the method we developed in our previous studies not valid. Thus, we have to redesign the feature extraction method. We choose 1~25 Hz instead α band in tradition. By using all the EEG data between 1~25 Hz, the computational loading of OMAP becomes very heavy. So we use PCA to reduce the data dimensions.

4 Results

The multi-task scheduling mechanism was developed to ensure the accurate sampling rate for EEG signal acquisition and data process and analysis in real time. In order to test the performance of the embedded multi-task scheduling system, we use two-channel data (sampling rate is 65Hz) EEG signals. The BCI system executes the signal processing procedure in DSP for 1000 times for the comparison of with and without embedded multi-task scheduling system performance. It takes 2425 seconds for a system without multi-task scheduling to complete the 1000 cycles, whereas the system with multi-task scheduling needs only 1933 seconds. However, the execution time is mainly used in receiving data. It takes 1838 seconds for a BCI system to complete 1000 cycles for only receiving data, which means it takes only 587/95 seconds for the BCI system without/with multi-task scheduling to complete 1000 cycles. The system without embedded multi-task scheduling needs 2.425 seconds to complete one cycle, while the system with embedded multi-task scheduling needs only 1.933 seconds. If the time of receiving data is not considered, the executing time of system without embedded multi-task scheduling is 0.6 second per cycle, and the executing time will be less than 0.1 second if the system uses embedded multi-task scheduling. As a result, the embedded multi-task scheduling system can help to reduce execution time and ensure the accurate frequency of receiving data.



(a) Training result



(b) Testing result

Fig. 4. The training and testing results of driving performance

In order to test and verify that the analysis algorithm is reliable to estimate driver's drowsiness level, EEG data are collected and tested on both PC and the proposed system. Two different sessions of EEG signal are acquired in different day for each subject. The EEG data collected in the 1st session is used as training data to construct a driving error estimating system. The EEG data collected from the 2nd session is then applied to the constructed estimating system to predict the driving performance of the driver. The correlation coefficient between the predicted driving performance and the actual driving performance recorded in the second session is calculated and is considered as an index of system performance. The training and testing results of the driving performance estimation is shown in Fig. 4.

5 Conclusion

A wireless embedded BCI system with real-time bio-signals processing ability is proposed in this paper. It consists of a multi-channel physiological acquisition and amplification unit, a wireless transmission unit, a dual-core signal processing unit, a remote real-signal display and monitoring unit, and a warning device. The EEG signal was first acquired by signal acquisition and amplification unit, and then transmitted from wireless data transmitter to wireless data receiver. The wirelessly transmitted EEG signals are processed by the data processing unit and the processed results are transmitted to the remote system for data storage, real-time display or triggering the warning devices by TCP/IP. A multi-task scheduling procedure is employed in the dual-core signal processing unit to enhance the efficiency of the embedded system and make sure the BCI system can properly work in real-time. A real-time drowsiness detection method combined with an on-line warning feedback is implemented in the developed system for demonstration. The uniqueness of the implemented system includes: (1) real-time signal processing ability (2) selectable wireless transmission which provides wide variety of applications (3) the dual-core embedded system with multi-task scheduling procedure for concurrent signal processing and (4) novel dry electrodes that ensure the convenience of using this system.

References

1. Bianchi, L., Babiloni, F., Cincotti, F., Arrivas, M., Bollero, P., Marciani, M.G.: Developing Wearable Bio-Feedback Systems: A General-Purpose Platform. IEEE Transaction on Neural Systems and Rehabilitation Engineering 11, 117–119 (2003)
2. Hsieh, H.Y., Liang, S.F., Ko, L.W., Lin, M., Lin, C.T.: Development of a Real-Time Wireless Embedded Brain Signal Acquisition/Processing System and its Application on Driver’s Drowsiness Estimation, In: Proc. of the 2006 IEEE International Conference on Systems, Man, and Cybernetics (October 8-11, 2006)
3. Kobayashi, T., Takahashi, K.: Linux DSP Gateway Specification Rev2.0, Nokia Corporation (November 13, 2003)
4. Jane, W.S.L.: Real-Time Systems. Prentice-Hall, Englewood Cliffs (2000)
5. Lin, C.T., Wu, R.C., Jung, T.P., Liang, S.F., Huang, T.Y.: Estimating alertness level based on EEG spectrum analysis. EURASIP Journal on Applied Signal Processing 2005(19), 3165–3174 (2005)
6. Lin?, C.T., Wu, R.C., Liang, S.F., Chao, W.H., Chen, Y.J., Jung, T.P.: EEG-based drowsiness estimation for safety driving using independent component analysis. IEEE Transactions on Circuits and Systems I 52(12), 2726–2738 (2005)

Modelling Cognitive and Affective Load for the Design of Human-Machine Collaboration

Mark A. Neerincx

TNO Human Factors, P.O. Box 23, 3769 ZG, Soesterberg, The Netherlands,
Delft University of Technology,
4, Mekelweg, 2628 CD, The Netherlands
mark.neerincx@tno.nl

Abstract. We are developing models for hybrid human-machine systems that can cope autonomously with unexpected, complex and potentially hazardous situations. The synthetic or electronic partner (*e*Partner) has to acquire and maintain knowledge of the (momentary) cognitive and affective load of the tasks and situation, and the capacities of the human partner (*h*Partner) to cope with this load. For adequate partnership, cognitive and affective load models are needed that support shared situation awareness, trust and scrutability. This paper presents two such models that are being developed and tested for military and space operations in situated cognitive engineering cycles.

Keywords: mental load, emotion, human-machine collaboration, synthetic or electronic partner, and cognitive engineering.

1 Introduction

Technological developments, e.g., on ambient intelligence and context-aware services, enable the design of joint cognitive systems in which the human and machine actors collaborate in an effective and efficient way. Such systems provide new possibilities to cope autonomously with unexpected, complex and potentially hazardous situations by mutual human-machine amplification of individual capabilities and by combining human and machine cognitive resources for situation assessment, problem solving and planning. Specifically, we aim at the design of a collection of distributed and connected personal, synthetic or electronic, partners (*e*Partners) to support the human partners (*h*Partners) in the military, space and medical domain. For these three domains, important goals of the partnership are, respectively, to dynamically attune the task allocation and level of automation to the available cognitive capacities and work context of operations [1], to improve human-machine team's resilience and safeguard *h*Partners from failures [2], and to improve the self-care of chronic patients [3].

1.1 *e*Partner

To establish the goals of dynamic task allocation, team resilience or patient's self-care, the *e*Partner has to acquire and maintain knowledge of the (momentary)

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 568–574, 2007.

cognitive and affective load of the tasks and situation, the capacities of its *h*Partner to cope with this load, and *h*Partner's intentions. In general, an *e*Partner has knowledge of its *h*Partner with respect to his or her permanent characteristics (e.g., personality), dynamic characteristics (e.g., experience), base-line state (e.g., "normal" heart rate), momentary state (e.g., current momentary heart rate), and tasks (e.g., alarm handling). Based on this knowledge, the *e*Partner maintains a model of the task demands that are critical for its *h*Partner (e.g., the risks of cognitive lock-up in complex task situations; [4]). It will have different mitigation strategies to prevent or to diminish negative effects of human operations in such critical situations by taking over some tasks, guiding the task performance, requesting other partners to help, or subtle actions to keep the human in an adequate state (e.g. open-mindedness, alertness).

The knowledge or models that *e*Partners maintain of their *h*Partners should support the sharing of knowledge and maintenance of an adequate trust level.

1.2 Shared Knowledge

The *e*Partner should be able to express and share its knowledge, and to express its capabilities to apply this knowledge for the collaborative activities. Partner's expressions of their cognitive capacities and emotions are crucial for real collaboration, for example, for effective critiquing [5] or persuasion [6]. The user interface of the *e*Partner is "natural or intuitive" by expressing and interpreting communicative acts based on a common reference of the human and machine actors.

A shared understanding of the current situation and the resources that are available for the required activities is needed for collaboration. It is important that the *h*Partner can access *e*Partner's knowledge about the situation and him or her, and that he or she has the possibility to correct or add *h*Partner's knowledge. He or she needs to know what the "*e*Partner knows about him or her", setting requirement for the scrutability of the models [7]. The humans should be able to inspect and control the details of the information held about them and the context in which they operate, the processes used to gather the information and the way that it is used. It may be possible to change some values according to his or her view (or according to the view of another partner of the team).

1.3 Trust

To really collaborate with a "knowledgeable" *e*Partner, the *h*Partner must trust it. Given the dependency of the astronauts on MECA and the ways the human-machine collaboration will be shaped, a high level of trust is required. For trust, we distinguish four dimensions: the experience, the persistence and competence of system behavior, the perceived servitude of the system, and the understanding of the system's content and operations. Trust in automation has both cognitive aspects, expressed in beliefs and expectations about the automation, as well as affective and motivational aspects, expressed in feelings and intentions toward the automation [8,9]. Sharing knowledge as described in section 1.2 is expected to support trust.

In sum, the ePartners must have knowledge of the momentary cognitive and affective load the tasks and contexts bring about for each team member. Furthermore, they should be able to communicate this knowledge with the human team members.

Therefore, we develop and apply so-called practical or "simple" theories on cognitive and affective load. Such a theory has face validity and comprises accepted features of human cognition, to be "contextualized, quantified and instantiated" for the application domain such as defense and space missions. Multimodal user-state, user-behavior and context sensing technology is used to "feed" the load models. The nest sections present two load models that are being developed and tested for military and space operations in situated cognitive engineering cycles.

2 Cognitive Task Load

Neerincx [10] developed a model of cognitive task load (CTL) and applied it for task allocation and the design of adaptive interfaces. This model could be part of the knowledge that the *e*Partner has of its *h*Partner, distinguishing three types of cognitive load factors.

First, the *e*Partner should have knowledge of the *time pressure*. In addition to the operational and contextual demands, human's cognitive processing speed determines this pressure for an important part, that is, the speed of executing elementary cognitive processes. Particularly, time pressure is high when the processes require a lot of attention and focused concentration (cf. [11]). Cognitive processing speed is determined by the individual capabilities to search and compare known visual symbols or patterns, to perform simple (decision-making) tasks, and to manipulate and deal with numbers in a fast and accurate way. Second, the *task complexity* affects the cognitive task load. Task information that is processed automatically, results into actions that are hardly cognitively demanding. Performance of routine procedures results into relatively efficient problem solving. Problem solving and action planning for relatively new situations can involve a heavy load on the limited capacity of working memory. Humans expertise and experience with the tasks have substantial effect on their performance and the amount of cognitive resources required for this performance. Higher expertise and experience result in more efficient, less-demanding deployment of the resources. Third, the CTL theory distinguishes *task switching or sharing* as a third load factor to address the demands of attention shifts or divergences. Complex task situations consist of several different tasks, with different goals. These tasks appeal to different sources of human knowledge and capacities and refer to different objects in the environment. Switching entails a change of applicable task knowledge.

The effects of cognitive task load depend on the concerning task duration (Table 1) In general, the negative effects of under- and overload increase over time. Under-load will only appear after a certain work period, whereas (momentary) overload can appear at every moment. When task load remains high for a longer period, carry-over effects can appear reducing the available resources or capacities for the required human information processing. Vigilance is a well-known problematic task for operators in which the problems increase in time. It can result in either stress due to the requirement to continuously pay attention on the task or boredom that appears with highly repetitive, homogeneous stimuli.

Table 1. Overview of 4 negative effects of cognitive task demands for a certain task period

	Task Performance Period		
	Short (<5min)	Medium (5-20min**)**	Long (>20min)
Time pressure *Low* Complexity *Low* Task switches *Low*	no problem	Under-load	
Time pressure *High* Complexity *Low* Task switches *Low*	no problem		Vigilance
Time pressure *High* Complexity *All* Task switches *High*	Cognitive lock-up		
Time pressure *High* Complexity *High* Task switches *High*	Overload		

3 Affective Load

Affection, emotion and mood are concepts that can have many interpretations. We will use affection and emotion interchangeably to reflect a momentary state, and mood to describe a can last for a state with a longer duration. Affection comprises a broad range of feelings that humans can have and which can influence humans in their behavior [12]. For characterizing the affective load, we focus on the underlying, often physiologically correlated factors (e.g. arousal) and map these onto distinct dimensions. Such dimensional models are helpful in both recognition and expression, as well as in models of emotion generation, in situations where sufficient data may not be available for more highly differentiated responses. Based on the Pleasure-Arousal-Dominance (PAD) model of Mehrabian, we distinguish two dimensions to

Table 2. The 2D model of affection or emotion with example effects

		Valence	
		NEGATIVE	**POSITIVE**
Arousal	**HIGH**	tunnel vision and higher concentration	divergent and creative thinking and problem solving
	LOW	boredom	relaxation

define the emotional state: the arousal level—low versus high—and the valence—positive versus negative (table 2). We do not distinguish a separate dominance dimension like the original PAD-model, because the dominance scale proved to explain the least variance and had the highest variability in terms of its inferred meaning in previous research.

Emotions can be measured through different modalities. Usually, physiological measures such as heart rate or skin conductivity are considered obtrusive, while speech and facial expressions are relatively non-obtrusive measures.

4 Cognitive Engineering for H-M Partnership

Due to the adaptive nature of both the human and machine behavior, it is difficult to provide generic and detailed predictions on the overall human-machine performance. Therefore, Neerincx & Lindenberg developed a situated cognitive engineering method [13]. First, the technological design space sets a focus in the process of specification and generation of ideas. Second, the reciprocal effects of technology and human factors are made explicit and are integrated in the development process. As shown in the previous sections, the human factors knowledge provides relevant theories, guidelines, support concepts and methods for the specification and assessment of H-M partnerships. In the specification, both the guidelines and the technological design space must be addressed concurrently. In the assessment it is checked whether the specifications agree with these guidelines and the technological design space.

Furthermore, the practical theories of cognitive and affective are being refined, situated and validated in the domain of application. For realizing adequate H-M partnership, generic human-factors knowledge and *e*Partner concepts are refined, contextualized and tested within the domain. The situated cognitive engineering framework has been developed and applied in the defense, space and medical domain to enhance the capacities of teams and team-members during critical and complex tasks (e.g., to improve task load management, trouble-shooting and situation awareness).

For example, for future manned space missions to the Moon or Mars, we specified a number of partnership scenarios. One scenario starts with two human-machine teams, team A and B, exploring the surface at different locations. Team B is working at a large distance from the habitat, and has a relatively large rover that can carry an astronaut. At the habitat, one astronaut is doing her exercises following her ‘self-care program’. For one member of team A, Charles, the spacesuit heater fails (figure 1). Team A, i.e. consisting of *e*Partners and astronauts, starts a fault detection and diagnosis process. The *e*Partner detects the *affective state* “panic”, predicts hypothermia and calls for help. In parallel, the following actions are started:

- Habitat prepares to receive astronaut (goal resetting)
- *e*Partner starts rescheduling the activities of actors, based on current *cognitive task load* states
- Rover in other team offers help & starts out
- *e*Partner informs astronaut (& others) of plan
- Astronaut faints earlier than predicted
- *e*Partner & rover devise way to pick up astronaut
- Rover transports astronaut to habitat

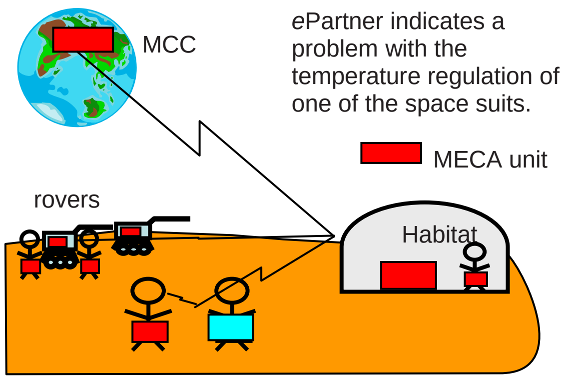


Fig. 1. Example scenario state for the suit failure (MCC = Mission Control Center)

Note that the focus is on the performance of the mental activities of human actors and the cognitive functions of machine actors, to achieve the (joint) operational goals. In this way, the notion of collaboration has been extended, viewing the machine as a social actor that can take initiative to act, critique or confirm in joint human-machine activities. The design focuses on the manifestation of these activities in "real settings", corresponding to the concept of macrocognition [14]. The cognitive engineering method is based on experiences with previous and current task performances (space, navy, care sector) and based on practical theories as described above.

5 Conclusions

We are developing models for hybrid human-machine systems that can cope autonomously with unexpected, complex and potentially hazardous situations. The synthetic or electronic partner (*e*Partner) has to acquire and maintain knowledge of the (momentary) cognitive and affective load of the tasks and situation, and the capacities of the human partner (*h*Partner) to cope with this load. For adequate partnership, cognitive and affective load models are needed that support the sharing of knowledge and acquisition of adequate trust levels. This paper presented two such models that are being developed and tested for military, space and medical operations, the models for cognitive and affective load. Test results are being used to improve the models and to implement them into *e*Partners.

Building an automatic cognitive and affective load recognition system can be very complex, especially if we want to incorporate an accurate and complete model or theory of cognition or affection. However, it may not always be necessary or realistic to pursue an ideal model; detection of 'simple' striking load states in context (e.g., 'panic') can also be of high practical value to realize effective partnership.

Acknowledgements. The development of *e*Partners for the Navy is funded by the Minister of Defence, for manned space missions by the European Space Agency

(Contract Number 19149/05/NL/JA), and for medical care by the Dutch IOP MMI program of SenterNovem.

References

1. Grootjen, M., Neerincx, M.A., van Weert, J.C.M.: Task-based interpretation of operator state information for adaptive support. In: Schmorrow, D.D., Stanney, K.M., Reeves, L.M. (eds.) Foudations of Augmented Cognition, 2nd edn. pp. 236–242. Strategic Analysis, Arlington, Virginia (2006)
2. Neerincx, M.A., Lindenberg, J., Smets, N., Grant, T., Bos, A., Olmedo Soler, A., Brauer, U., Wolff, M.: Cognitive Engineering for Long Duration Missions: Human-Machine Collaboration on the Moon and Mars. In: SMC-IT 2006: 2nd IEEE International Conference on Space Mission Challenges for Information Technology, pp. 40–46. IEEE Conference Publising Services, Los Alamitos, California (2006)
3. Looije, R., Cnossen, F., Neerincx, M.A.: Incorporating Guidelines for Health Assistance into a Socially Intelligent Robot. In: Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man 2006), pp. 515–520. University of Hertfordshire, Hatfield, UK (September 6-8, 2006)
4. Kerstholt, J.H., Passenier, P.O.: Fault management in supervisory control: the effect of false alarms and support. Ergonomics 43(9), 1371–1389 (2000)
5. Fischer, G.: User modeling in Human-Computer Interaction. User. Modeling and User.-Adapted Interaction 11, 65–68 (2001)
6. Fogg, B.J.: Persuasive technology: Using computers to change what we think and do. Morgan Kaufmann Publishers, Amsterdam etc (2003)
7. Kay, J., Kummerfeld, B., Lauder, P., Personis, A.: Server for User Models. In: De Bra, P., Brusilovsky, P., Conejo, R. (eds.) AH 2002. LNCS, vol. 2347, pp. 203–212. Springer, Heidelberg (2002)
8. Muir, B.: Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. Ergonomics 37, 1905–1922 (1994)
9. Lee, J.D., See, K.A.: Trust in automation: Designing for appropriate reliance. Human Factors 46, 50–80 (2004)
10. Neerincx, M.A.: Cognitive task load design: model, methods and examples. In: Hollnagel, E. (ed.) Handbook of Cognitive Task Design (Chapter 13), pp. 283–305. Lawrence Erlbaum Associates, Mahwah, NJ (2003)
11. Carroll, J.B.: Human Cognitive Abilities: A Survey of Factor Analytic Studies. Cambridge University Press, New York (1993)
12. Cowie, R., Schröder, M.: Piecing together the emotion jigsaw. In: Machine Learning for Multimodal Interaction, 305–317 (2005)
13. Neerincx, M.A., Lindenberg, J.: Situated cognitive engineering for complex task environments. In: Schraagen, J.M. (ed.) Natural Decision Making & Macrocognition. Ashley (will appear in 2007) (in press).
14. Klein, G., Ross, K.G., Moon, B., Klein, D.E., Hoffman, R., Hollnagel, E.: Macrocognition, IEEE Intelligent Systems, 81–85 (May/June 2003)

Event-Related Brain Potentials Corroborate Subjectively Optimal Delay in Computer Response to a User's Action

Hiroshi Nittono

Graduate School of Integrated Arts and Sciences, Hiroshima University
1-7-1 Kagamiyama,
Higashi-Hiroshima 739-8521, Japan
nittono@hiroshima-u.ac.jp

Abstract. This study examined how the presentation timing of a computer response to a user operation affects attention allocated to the stimulus. Event-related brain potentials were recorded in response to auditory stimuli that were presented after a single mouse button press with three different durations of delay: 0, 150, and 300 ms. The amplitude of the P300 component, which is assumed to reflect the amount of attentional resources allocated to the eliciting event, increased when the stimuli were presented 150 ms after button press, compared with when the same stimuli were presented immediately (0 ms) or 300 ms after button press. These results are consistent with a previous psychophysical finding that the insertion of a moderate delay interval can increase the user's sense of control, and suggest that ERPs can be used as an objective tool for assessing the state of attention in a particular interface design.

Keywords: psychophysiology, P300, attention, self-paced task, interface design.

1 Introduction

In human-computer interaction (HCI), users often receive auditory or visual feedback that indicates the computer acknowledged or responded to their operation. It is generally believed that overall performance and user satisfaction are better when the delay in computer response is shorter [1,2]. Guidelines for designing user interface software recommend that the computer should acknowledge data entry actions rapidly and that delays in displayed feedback should not exceed 200 ms for normal operation [3]. However, they did not suggest the lower limit. Few empirical studies have examined what length of delay is optimal for users, that is, most comfortable and maximizes the user's sense of control.

A previous psychophysical study showed that an immediate response from a computer after a single mouse button press was not always felt most comfortable [4]. Both a category-scaling and an adjustment method of assessment were used to determine the optimal delay with which the user felt most comfortable. As expected, a computer response was judged to be too slow when the delay was 200 ms or longer

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 575–581, 2007.

for auditory and 300 ms or longer for visual stimuli. On the other hand, when the delay was 50 ms or shorter, an auditory or a visual feedback response was judged as too fast to be a consequence of the action. These results indicate that the subjectively optimal delay in a confirming feedback is between 100 and 200 ms for a single, discrete button press. Although this response preference may be formed through daily experience with current personal computers, the results suggest that the insertion of a moderate delay interval can increase the user's sense of control in some cases.

The present study was aimed to examine how the processing of an auditory stimulus after a mouse button press varied as a function of the duration of delay. For this purpose, event-related brain potentials (ERPs) were expected to be an evaluation tool. ERPs are voltage fluctuations that are associated in time with the occurrence of some physical or mental event, and have been used in engineering psychology [5]. In particular, the amplitude of the P300 component is assumed to reflect the amount of attentional or processing resources allocated to the eliciting stimulus [6]. Nittono and colleagues have suggested that ERPs elicited by a computer response to a user operation can provide useful information about the state of attention in HCI tasks [7,8]. They showed that the amplitude of the P300 increased when the stimuli were presented after a participant's voluntary button press compared with when the same stimuli were presented at an external pace of a computer [9,10]. If the user feels more comfortable with the stimuli that are presented with a delay of 100−200 ms, more attentional resources would be allocated to the stimuli and a larger P300 would be elicited, compared with when the same stimuli are presented with a shorter or a longer delay.

2 Method

2.1 Participants

Sixteen right-handed university student volunteers (7 men, 9 women, M = 23.3 years old) participated in the study. All of them were right-handed and had normal hearing. They gave written informed consent.

2.2 Stimuli and Task

An auditory three-stimulus target detection (oddball) task was used. Participants were asked to press a button of a two-button computer mouse by an index finger once per 1 to 2 s. Each mouse click produced one of the three pure tones in a random order: standard (p = .70, 1000 Hz), target (p = .15, 2000 or 500 Hz), and nontarget tones (p = .15, 500 or 2000 Hz). The tone duration was 70 ms including rise and fall times of 10 ms each. They were presented through headphones at 60 dB SPL. The task involved button-press responses to target stimuli using the other index finger. The pitches of target and nontarget tones and the trigger and response fingers were counterbalanced across participants.

2.3 Procedure

In three separate conditions, a delay of 0 ms (too fast), 150 ms (optimal), or 300 ms (too slow) was inserted between a participant's button press and the onset of feedback tone. Each condition consisted of two blocks with 140 trials each. During the task, participants

were asked to look continuously at a fixation point on a CRT display in front of them. After each condition, participants rated how they felt the timing of stimulus presentation (−5: fast, +5: slow) and the difficulty of the task (−5: easy, +5: difficult) on 11-point scales. The order of conditions were counterbalanced across participants.

2.4 Physiological Recording

An electroencephalogram (EEG) was recorded from 51 scalp sites according to the 10 % system using an elastic electrode cap (Quik-Cap, Compumedics USA, TX). A high-pass filter of 0.032 Hz (time constant 5 s) and a low-pass filter of 60 Hz were used for recording. The sampling rate was 500 Hz. The data were re-referenced to digitally linked mastoids offline. Horizontal and vertical electrooculograms (EOGs) were recorded from the outer canti of both eyes and from above and below the left eye, respectively.

2.5 Data Reduction

ERPs were calculated separately for each stimulus type and condition by averaging the 1,000-ms period starting from 200 ms before the stimulus onset. Trials in which the EEG or EOGs exceeded ±100 μV were excluded from the ERP analysis. The first 200-ms period of each waveform served as the baseline. To cancel out the movement-related potentials associated with a trigger button press, difference waveforms were calculated by subtracting the ERPs to standard stimuli from the ERPs to target and nontarget stimuli in each condition [9,10]. The P300 was designated as a positive wave between 250 and 500 ms after the onset of target and nontarget stimuli. Its peak was identified in the difference waveforms at the parietal midline site, Pz, where the P300 usually shows the maximal amplitude. Then, P300 amplitudes at three midline sites (frontal: Fz, central: Cz, parietal: Pz) were measured at this peak latency.

2.6 Statistical Analysis

Subjective, behavioral, and P300 amplitude and latency data were submitted to analyses of variance (ANOVAs) with repeated measures. To control the type I error associated with violation of the sphericity assumption, degrees of freedom greater than one were adjusted by the Greenhouse-Geisser ε correction. Post hoc comparison was made by using Shaffer's modified sequentially rejective multiple test procedure, which extends Bonferroni *t* tests to a stepwise fashion [11]. The significance level was set at .05 for all statistical tests.

3 Results

3.1 Subjective and Behavioral Measures

Table 1 summarizes subjective ratings and behavioral data. The results of one-way ANOVAs with a factor of condition are shown along with the results of post hoc comparisons. Participants judged the stimuli to be more well-timed in the 150-ms delay condition than in the 0-ms or 300-ms conditions. The task was rated easiest in the 150-ms delay condition, in which the difficulty rating was significantly lower than

in the 300-ms delay condition. Mean button press interval increased from the 0-ms delay condition to the 300-ms delay condition. Reaction times in response to target tones were longer in the 0-ms delay condition than in the 150-ms and 300-ms conditions, the latter of which did not differ significantly from each other. The duration of delay did not affect error rates significantly.

Table 1. Means ± standard errors of subjective and behavioral measures

Measures	Delay: 0 ms	150 ms	300 ms	p
Subjective rating				
Timing (–5: *fast* – +5: *slow*)	–0.6 ± 0.4[a,b]	0.1 ± 0.3[b,c]	1.4 ± 0.5[a,c]	.004
Difficulty (–5: *easy* – +5: *difficult*)	–0.4 ± 0.5	–1.2 ± 0.6[b]	0.5 ± 0.6[a]	.015
Mean button press interval (ms)	1,394 ± 43[b]	1,462 ± 43[b]	1,554 ± 34[a,c]	.003
Mean reaction time (ms)	490 ± 21[a,b]	455 ± 18[c]	451 ± 19[c]	.004
Miss (%)				
Target	2.4 ± 0.5	1.6 ± 0.6	2.4 ± 1.0	.536
False alarm (%)				
Standard	0.3 ± 0.1	0.5 ± 0.2	0.3 ± 0.1	.397
Deviant	0.4 ± 0.2	0.4 ± 0.2	1.3 ± 0.7	.167

[a]Significantly different from the value in the 150-ms condition.
[b]Significantly different from the value in the 300-ms condition.
[c]Significantly different from the value in the 0-ms condition.

3.2 ERPs

Fig. 1 shows grand mean ERP waveforms averaged across all participants. Target and nontarget stimuli elicited a P300 wave in all conditions. Fig. 2 shows difference waveforms in which movement-related potentials were canceled out. The amplitude of the P300 appears to be largest in the 150-ms delay condition. Fig. 3 shows the peak amplitude measurements of the P300 in each condition and site. A three-way ANOVA with factors of condition, stimulus, and site showed a significant main effect of condition, $F(2, 30) = 4.05$, $p = .034$, $\varepsilon = .882$. No interaction effects including the factor of condition were found. Post hoc comparisons were made on the amplitude values averaged across stimulus types and sites. P300 amplitude was significantly larger in the 150-ms condition ($M = 7.3$ µV) than in the 0-ms and 300-ms conditions (Ms = 5.9 and 5.7 µV, respectively). The main effect of stimulus was also significant, $F(1, 15) = 48.59$, $p < .001$, which suggests that target stimuli elicited a larger P300 than did nontarget stimuli. Moreover, the main effect of site and interaction of stimulus and site were significant, Fs(2, 30) = 41.57 and 29.90, ps < .001, εs = .659 and .613. Post hoc analyses showed that the P300 in response to target stimuli had a parietal scalp distribution (Fz < Cz < Pz), whereas the P300 in response to nontarget stimuli had a centro-parietal scalp distribution (Fz < Cz = Pz, no significant difference in amplitude between Cz and Pz).

P300 latency was shorter for nontarget stimuli ($M = 329$ ms) than for target stimuli ($M = 377$ ms), but did not vary significantly across conditions. A two-way ANOVA with factors of condition and stimulus showed a significant main effect of stimulus, $F(1, 15) = 15.19$, $p <. 001$, but no significant main and interaction effects of condition.

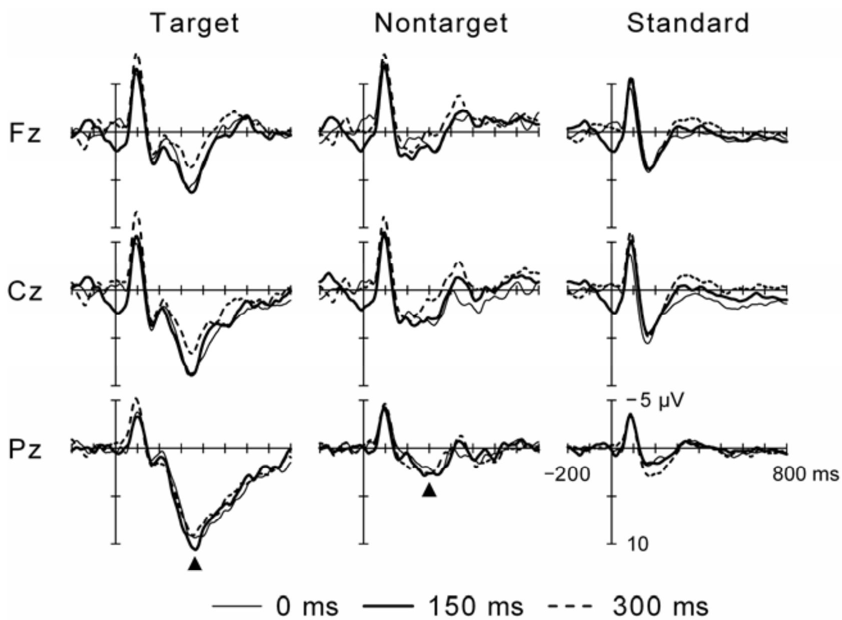


Fig. 1. Grand mean ERP waveforms elicited by target, nontarget, and standard tones presented with three different durations of delay (N = 16). Solid triangles indicate the P300. Fz: frontal midline electrode site, Cz: central midline electrode site, Pz: parietal midline electrode site.

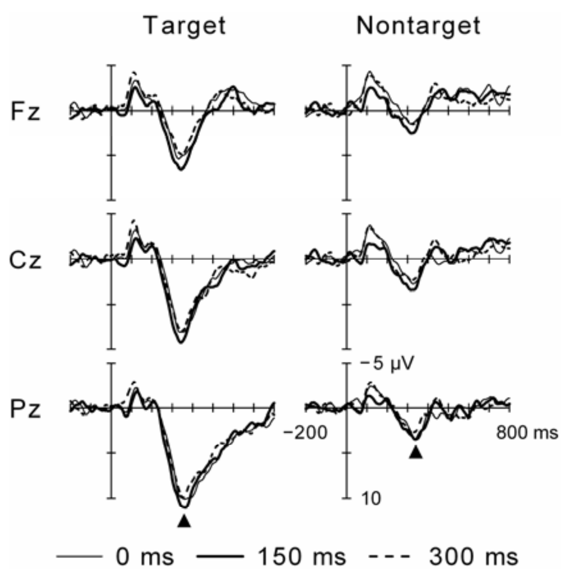


Fig. 2. Difference waveforms calculated by subtracting the ERPs elicited by standard tones from the ERPs elicited by target and nontarget tones (N = 16). Solid triangles indicate the P300.

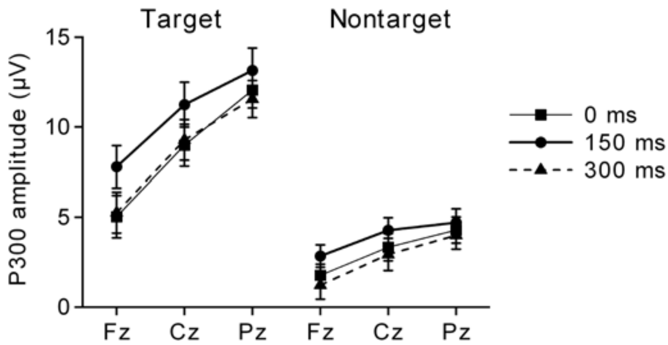


Fig. 3. The peak amplitude measurements of the P300 in each condition

4 Discussion

In an auditory target detection task, participants felt the stimuli more well-timed and the task easier when the stimuli were presented after a voluntary button press with a delay of 150 ms, compared with when the same stimuli were presented immediately or 300 ms after a button press. These results are consistent with the previous psychophysical finding that the subjectively optimal delay in a computer response to a user's single, discrete button press is between 100 and 200 ms.

Both target and nontarget stimuli elicited a P300 wave, the amplitude of which was larger for target stimuli. This result agrees with the literature in that P300 amplitude is determined additively by stimulus probability and task relevance [12]. While these variables were constant, P300 amplitude increased in the 150-ms delay condition. According to the prevailing view about the P300 [5,6], this result suggests that participants allocated more attentional resources to the stimuli in this condition.

Behavioral measures did not show any advantageous effect of the 150-ms delay. Reaction times were significantly longer in the 0-ms delay condition, possibly because of the motor conflict between the right and left index fingers that were used for triggering and responding. Mean button press interval increased along with increasing delays between button press and stimulus presentation. This result agrees with the finding that interresponse times increased when the print mechanism of a teletype was delayed in relation to typing on the keyboard [1]. It is known that P300 amplitude tends to be larger when interstimulus intervals were longer [13]. However, the P300 amplitude differences in the present study are not attributable to the differences in button press intervals, because the middle length produced the largest amplitude.

5 Implications for Application

To my knowledge, this is the first empirical study to use ERPs as an objective tool for assessing human-computer interface designs. Although the 150-ms delay was found to be optimal in this study, the exact figure is not critical nor universal, because it was derived from a special case, i.e., a target detection task in which the stimuli were

presented after a single button press. The optimal delay for multiple key stroking may be shorter than 150 ms [14]. The present study is not intended to propose a new standard of delay interval, but to demonstrate that ERPs can be recorded in HCI tasks and provide useful information that corroborates or complements subjective and behavioral measures. In particular, the ERP technique is advantageous when researchers want to know how much attention is allocated to a particular stimulus or event.

Acknowledgments. The author thanks to Koichi Shimizu and Tadao Hori for their help in conducting experiments. This research was supported by a Grant-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science, and Technology (No. 18730466).

References

1. Long, J.: Effects of delayed irregular feedback on unskilled and skilled keying performance. Ergonomics 19, 183–202 (1976)
2. Williges, R.C., Williges, B.H.: Modeling the human operator in computer-based data entry. Hum. Factors 24, 285–299 (1982)
3. Smith, S.L., Mosier, J.N.: Guidelines for designing user interface software (ESD-TR-86-278). The MITRE Corporation, Bedford, MA (1986)
4. Nittono, H., Shimizu, K., Hori, T.: Subjectively optimal delay in computer response to a user's voluntary action. Percept. Mot. Skills 99, 924–930 (2004)
5. Donchin, E., Kramer, A.F., Wickens, C.D.: Applications of brain event-related potentials to problems in engineering psychology. In: Coles, M.G.H., Donchin, E., Porges, S.W. (eds.) Psychophysiology: Systems, Processes, and Applications, pp. 702–718. Guilford, New York (1986)
6. Kok, A.: On the utility of P3 amplitude as a measure of processing capacity. Psychophysiology 38, 557–577 (2001)
7. Nittono, H., Hamada, A., Hori, T.: Brain potentials after clicking a mouse: A new psychophysiological approach to human-computer interaction. Hum. Factors 45, 591–599 (2003)
8. Nittono, H.: Missing-stimulus potentials associated with a disruption of human-computer interaction. Psychologia 48, 93–101 (2005)
9. Nittono, H.: The action-perception paradigm: A new perspective in cognitive neuroscience. Int. Congr. Ser. 1270, 26–31 (2004)
10. Nittono, H.: Voluntary stimulus production enhances deviance processing in the brain. Int. J. Psychophysiol 59, 15–21 (2006)
11. Shaffer, J.P.: Modified sequentially rejective multiple test procedures. J. Am. Stat. Assoc. 81, 826–831 (1986)
12. Johnson, R.: The amplitude of the P300 component of the event-related potential: Review and synthesis. In: Ackles, P.K., Jennings, J.R., Coles, M.G.H. (eds.) Advances in psychophysiology. JAI, Greenwich, CT, vol. 3, pp. 69–138 (1988)
13. Gonsalvez, C.J., Polich, J.: P300 amplitude is determined by target-to-target interval. Psychophysiology 39, 388–396 (2002)
14. Boyle, J.M., Lanzetta, T.M.: The perception of display delays during single and multiple keystroking. In: Proceedings of the Human Factors Society, vol. 28, pp. 263–266 (1984)

Effects of Pattern Complexity on Information Integration: Evidence from Eye Movements

Yanju Ren[1,2], Yuming Xuan[1], and Xiaolan Fu[1,*]

[1] State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
[2] Graduate School of the Chinese Academy of Sciences, Beijing, 100049, China
{renyj, xuanym, fuxl}@psych.ac.cn

Abstract. The present study employed empty cell localization paradigm and eye-tracking method to investigate the effects of memorized stimuli complexity on information integration between in visual short-term memory (VSTM) and visual perception. Two arrays of dots were displayed in sequence within a grid. Between the two arrays, one cell was always empty, and the participants' task was to specify the location of this "missing dot" It was found that the accuracy decreased as dot pattern of array 1 increased in complexity, especially under long ISI condition. The analysis of eye movement behavior, especially fixation location, demonstrated that participants were more likely to try to remember the location of the empty cells of array 1 other than locations of dots. From aspect of eye movement, these results offered the first evidence supporting convert-and-compare hypothesis.

Keywords: information integration, visual short-term memory, visual perception, eye movement.

1 Introduction

Visual short-term memory (VSTM) allows the contents of visual perception to be retained momentarily to create temporal continuity in a constantly changing visual environment [12] [18], yet it has severe capacity limits. Only four objects or six spatial locations can be retained in VSTM [14] [17]. Research effort in the past decade has focused primarily on the representation of a single visual display. Yet, in many everyday activities, visual information not only occupies space but also evolves over time. In order to keep coherent visual environment, we need to integrate the information retained in the VSTM and visual percept. Driving requires the constant updating of information gathered from many instances. When interacting with computer we need to integrate information from different windows.

Recently, empty cell localization paradigm, used to investigate perceptual integration [5] [6] [7] [9] was also used to examine the integration between VSTM

* Corresponding author.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 582–590, 2007.

and visual perception [1] [2] [3] [4] [10] [13]. In this paradigm, two arrays of dots were displayed in sequence within a visible or virtual grid. The first array filled approximately half of the cells in the grid. Array 2 filled all but one of the cells not filled in Array 1. Between the two arrays, one cell was always empty, and the task was to specify the location of this “missing dot”.

To date, two hypotheses were proposed for the mechanism of integration between VSTM and visual perception. One was image-percept integration hypothesis, which assumed visual perception could be integrated with the contents of VSTM [1] [2] [3] [4]. When an observer generated a representation of array 1 in memory, a subsequently perceived array 2 could be directly incorporated into the existing representation. The other was convert-and-compare hypothesis, which assumed the first array might be used to visually mark grid positions as locations that cannot constitute the correct answer [21] [22]. As a result, these positions might be inhibited, and attention directed to the positions that Array 1 left empty, enhancing the detection of the space left unfilled by the second array [10] [13].

The convert-and-compare hypothesis postulated that VSTM had a small capacity of about six spatial locations, and this hypothesis was consistent with limited capacity of VSTM [14] [17]. However, the image-percept integration hypothesis postulated that VSTM has a higher capacity of about 14-20 locations, and that it supports integration across sequential arrays [2] [4]. Given this hypothesis was correct, then what accounted for this apparently high-capacity memory, and could these data be reconciled with current views of limited-capacity VSTM? Hollingworth et al (2005) pointed out that it is likely that long-ISI performance was supported by VSTM, and VSTM grouped individual dots into a larger scale object or objects [10]. They tested this figural grouping hypothesis and found that pattern complexity had no effect on empty cell localization at 0-msec ISI, but there was a large simple pattern advantage at long ISIs.

In the present experiment, we continued to use Hollingworth et al's method to test further this figural grouping hypothesis [10]. More importantly, we monitored participants' eye movement behavior during integration in order to determine the locus of overt attention and to uncover the mechanism underlying the integration process. If the image-percept integration hypothesis is correct, participants should pay more attention to locations occupied by dots when viewing array 1; On the contrary, if the convert-and-compare hypothesis is correct, participants should pay more attention to empty cells instead.

2 Method

2.1 Participants

Twelve undergraduates (6 male and 6 female, their average ages were 21) from the China Agricultural University participated in the experiment. All participants reported normal or corrected-to-normal vision. After the experiment, they were paid 25 RMB.

2.2 Stimuli

The stimuli were similar to those used in Hollingworth et al (2005, Experiment 2) [10]. Two dot arrays were displayed on each trial within a 4 x 4 grid. The Array 1

stimuli were drawn from three sets: a simple figure set, a medium figure set and a complex figure set. Simple, medium and complex figures were generated based on complexity ratings compiled by Ichikawa [11]. Each stimulus collected by Ichikawa was a 4 x 4 grid with eight of the cells filled by dots. A total of 140 different patterns were rated by participants, and the stimuli were ordered from least complex (item 1) to most complex (item 140). The set of 32 simple, medium, and complex figure stimuli for the experiment were items 1-32, 55-86, and 109-140 from Ichikawa respectively. Sample stimuli from the simple, medium and complex figure set were showed in Fig. 1. The Array 2 stimuli were constructed by randomly filling seven of the eight cells not filled in Array 1. As a result, the position of the empty cell was also randomly determined. The 4 x 4 grid was composed of light-blue lines superimposed over the background (was light gray and subtended 36° x 27° of visual angle). Dots were presented in black. The entire grid subtended 20° of visual angle (both horizontally and vertically). Each cell in the grid subtended 5°. The diameter of each dot was 4°.

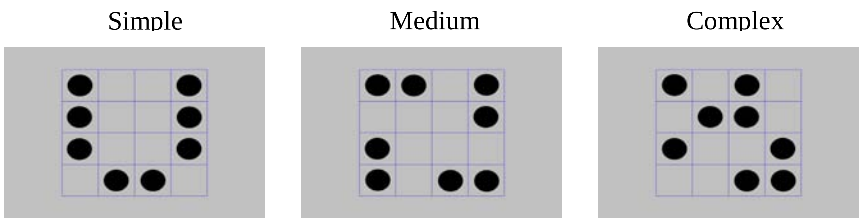


Fig. 1. Sample Array 1 stimulus from the simple pattern set (left), from the medium pattern set (middle) and from the complex pattern set (right) in the present experiment

2.3 Apparatus

The stimuli were presented at a resolution of 1024 by 768 pixels on a 19-inch video monitor at a refresh rate of 85 Hz. The presentation of stimuli and collection of responses was controlled by Experiment Builder software running on a Pentium IV PC. Viewing distance was maintained at about 60 cm. Eye position was sampled at a rate of 500Hz (every 2 ms) with an EyeLink II eyetracker (SR Research Ltd. in Canada) with a resolution of 0.01° in pupil only mode. The eyetracker and display monitor were interfaced with a computer that controlled the experiment. This system used video-based infrared oculography to measure eye and head position.

2.4 Design and Procedure

In general, the design and procedure were very similar to that used by Hollingworth et al (2005) [10]. On each trial, two dot arrays (eight and seven dots, respectively) were presented sequentially within an enclosed square grid separated by a variable ISI. There were three blocks of trials totally, with each level of pattern complexity (simple, medium and complex) comprising a block. On any given trial, one cell within the grid was never filled and the participants were instructed to identify the position of the empty cell.

The procedure was illustrated in Fig. 2 and each trial consisted of the following events. Firstly, the black point was presented in the center of the screen. When ready to begin, participants fixated the black point and pressed '5' on the Eye Link Button Controller to start the trial. There was the empty grid (superimposed over the gray background) which was presented for 500 ms delay before presentation of Array1. Array 1 was then presented within the grid for 35 ms (about three refresh cycles at 85 Hz). The variable ISI between the offset of Array 1 and the onset of Array 2 was 100, 750, 1500, or 2500 ms. During the ISI, the blank grid was displayed. Following the ISI, Array 2 was presented within the grid for 35 ms. Array 2 was followed by the grid with the 16 numbers from '1' to '16' presented in the center of each cell. Participants identified the location of the empty cell by speaking out the number standing for the cell. Then participants pressed '5' on the Eye Link Button Controller to start the next trial. Participants were asked to respond as accurately as possible and that they were under no speed stress.

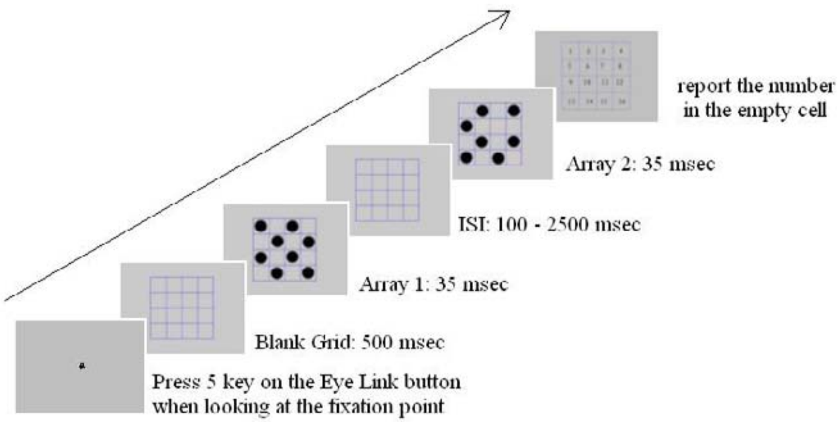


Fig. 2. Sequence of events in a trial of the experiment. The participants pressed 5 key on the Eye Link button to begin the trial, followed by the events illustrated in the figure. When the final blank grid with number appeared, the participants should speak out the number in the empty cell.

Participants were tested individually. Before tested, each participant was given a written description of the experiment along with a set of instructions. The experiment would get started after each participant understood the instructions.

The experiment consisted of two sessions: practice session and formal experiment session. The procedure of practice session was compiled by E-prime 1.1 and was similar to the formal experiment session except that there was no the black point screen and participant moved mouse to click the empty cell. Participants first completed a practice session of 16 trials, four in each of 4 ISI conditions, randomly intermixed. Feedback was provided in the practice session with the 1000 ms gray screen between trials either contained the word “correct” or “incorrect”. During the

practice session, eye movements of participants were not recorded. The experimental session, consisted of three blocks (simple pattern, medium pattern and complex pattern) of trials. Each block contained 128 trials and stimuli (32)-ISI (4) assignments were rotated within blocks by Latin square. The blocks' order was counterbalanced by Latin square between participants. Participants began each trial by fixating a center point. The position of the dominant eye was tracked, though viewing was binocular. At the beginning of each block, the eyetracker was calibrated and validated. Once calibration and validation were completed, the experimental trials began.

The eye tracker monitored eye position during the presentation of Array 1, ISI, and the presentation of Array 2. The participants began each trial by looking at the center of the grid but were free to move their eye without constraints during the trial. Eye position was scored in terms of the grid space that was fixated. Saccades were operationally defined as a change in fixation that exceeded 0.15° of visual angle accompanied by a velocity exceeding 30°/sec or an acceleration exceeding 8000°/sec^2 that was maintained for a minimum of 2 ms. The saccade was considered to be terminated when there criteria were no longer met.

The participants were not informed of the pattern complexity manipulation. Feedback was not provided in the experimental session. There were two-minute break between practice session and experimental session and ten-minute breaks between successive blocks. The entire experiment lasted approximately 90 min.

3 Results

The results were reported in two parts. First, general accuracy and error rates for different error types in the information integration task were reported. Second, various aspects of eye movement behavior were examined. Variables of interest included fixation number, fixation location, fixation duration and pupil size.

3.1 Integration Task Accuracy and Error Rate

A response was classified as correct, an Array 1 error (erroneously selecting a position occupied by the first array), or an Array 2 error (erroneously selecting a position occupied by the second array) and was measured in terms of the percentage of trials on which they occurred. The accuracy, percentage of Array 1 error and percentage of Array 2 error in the empty cell localization task was examined as a function of pattern complexity of Array 1 and ISI.

Accuracy. There were a reliable main effect of complexity [$F(2, 22) = 70.159$, $p < .001$], a reliable main effect of ISI [$F(3, 33) = 52.565$, $p < .001$] and a reliable interaction between complexity and ISI [$F(6, 66) = 3.714$, $p < .005$].

Array 1 Error. There was a reliable main effect of complexity [$F(2, 22) = 49.081$, $p < .001$], a reliable main effect of ISI [$F(3, 33) = 60.369$, $p < .001$] and a reliable interaction between complexity and ISI [$F(6, 66) = 3.170$, $p < .01$]. These data for accuracy and percentage of Array 1 error were illustrated in Fig. 3.

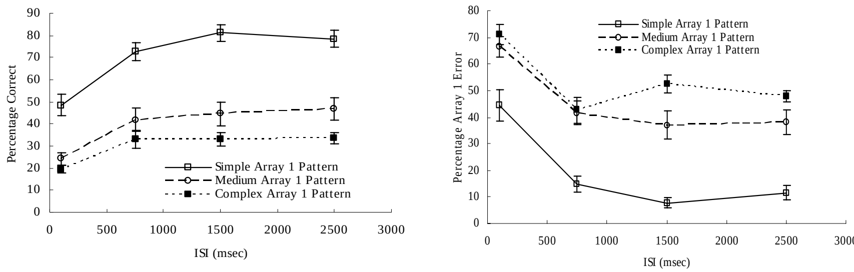


Fig. 3. Localization accuracy (the left) and array 1 error (the right) as a function of ISI and pattern complexity in present experiment. Error bars represent standard errors of the means.

Array 2 Error. Although there were a reliable main effect of complexity [$F(2, 22) = 8.856$, $p < .005$], a reliable main effect of ISI [$F(3, 33) = 5.095$, $p < .01$] and a reliable interaction between complexity and ISI [$F(6, 66) = 2.310$, $p < .05$], array 2 error is less than array 1 error in all conditions.

3.2 Eye Movement Behavior

To some extent, the locus of attention is indicated by the location in space that is fixated. Thus an analysis of modes in the distribution of fixations during the ISI separating the arrays can give insight into the manner in which attention is used during VSTM consolidation. In this section, fixation number, total fixation duration, average duration, fixation location and average pupil size were analyzed.

Fixation Number. ANOVA showed that the effects of different pattern complexities on fixation number were the same and there was a significant main effect of ISI [$F(3, 33) = 21.803$, $p < .001$]. The Pearson correlation between ISI and number of fixations was .99. As Fig. 4. (the left) illustrated, a linear regression analysis indicated that one additional fixation was made for every 1,000-msec increase in ISI.

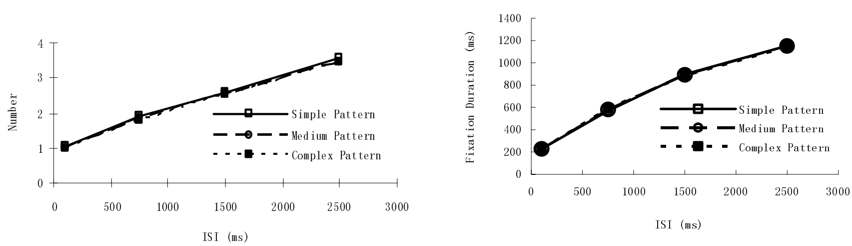


Fig. 4. The average fixation number (the left) and fixation duration (the right) in a trial in the experiment

Total Fixation Duration. Average total fixation durations on each trial were illustrated in Fig. 4. (the right) as a function of pattern complexity and ISI. An overall main effect of ISI was observed, since fixation durations generally increased with increases in ISI. ANOVA revealed that the effects of different pattern complexities on fixation

duration were the same and there was a significant main effect of ISI [$F(3, 33) = 36.886$, $p < .001$].

Average Fixation Duration. Average fixation duration on each trial across ISIs was 666.8 ms, 684.7 ms and 664.7 ms respectively under simple, medium and complex pattern conditions. ANOVA showed that the differences of average fixation duration among three pattern complexity conditions were not significant [$F < 1$, $p > .05$].

Fixation Location. Fixation location under the conditions of different ISIs and pattern complexities was illustrated in Fig. 5 (the left). The data revealed that the percentage of fixations on Array 1 dots was lower than chance level (50%) at any ISIs. ANOVA showed that there was a reliable main effect of pattern complexity [$F(3, 33) = 12.753$, $p < .001$] and a reliable interaction between pattern complexity and ISI [$F(6, 66) = 3.918$, $p < .005$]. Average percentage of fixation on dot locations across ISIs were 42.12%, 37.77%, and 44.39% respectively under simple, medium and complex pattern conditions, which were all below chance level [$| ts(11) | > 4$, $ps < 0.01$].

Pupil Size. Pupil size, measured by pupil diameter, under the condition of different ISIs and pattern complexities was illustrated in Fig. 5. (the right) ANOVA revealed that there was a reliable main effect of ISI [$F(3, 33) = 61.278$, $p < .001$].

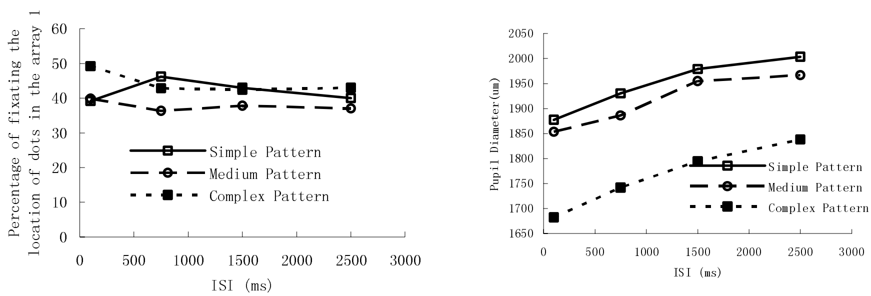


Fig. 5. The percentage of fixations on location originally occupied by a dot from Array 1 (the left) and average pupil diameter under the condition of different ISIs and pattern complexity (the right) in the experiment

4 Discussion

The effect stimuli complexity on information integration between VSTM and visual perception was studied in the present study. It was found that the accuracy in the simple pattern condition was higher than that in the medium pattern condition, which was higher than that in the complex condition, especially at long ISIs. Meanwhile, the array 1 error and the array 2 error also showed the similar trends. These findings supported the figural grouping hypothesis in VSTM and suggested that performance at long ISIs was supported by a limited-capacity, object-based visual memory that is sensitive to higher order image structure [10]. The effect of pattern complexity provided a means to reconcile apparently high-capacity memory at long ISIs with evidence that VSTM has a limited capacity of three or four objects. At long ISIs,

Array 1 is likely represented in VSTM as one or more higher order objects, each containing information from more than one array element.

Average fixation numbers on each trial were positively correlated with ISI, since longer ISI afforded more time for participants to make more eye movements. This finding showed that participants had eye movements during ISI. Fixation duration had been considered to reflect the amount of processing [8] [15] [19]. In the present study, Average total fixation duration observed on each trial was positively correlated with ISI, and average fixation duration under three pattern conditons were almost the same. Altogether, eye movements seemed to be a way which helped participants to retain information of array 1 in VSTM.

More importantly, fixation location analyses showed that the percentage of fixations on locations originally occupied by dots in Array 1 was less than chance level (0.50) in three pattern conditions. This finding suggested that the participants were more likely to pay attention to empty cell locations other than locations originally occupied by dots to retain array 1. Given that average duration of each fixation was almost the same under three pattern conditions, the total duration participants fixating locations of empty cells in array 1 was longer than total duration the participants fixating dots in array 1. These findings suggested that participants were more likely to try to retain locations of empty cells than locations of dots in array 1 during ISI. From aspect of eye movements, these results offer the first evidences supporting convert-and-compare hypothesis [13]. Brockmole et al (2005) also minitored the participants' eye movement and found that the percentage that participants fixate the grid position occupied by dots of array 1 was under random level [1]. The reasons for the difference between their results and the present study were not clear yet. One possible explanation could be the manipulation of pattern complexity in the present study. In Brockmole et al's study, the dot position in array 1 was randomly distributed, thus the dot pattern of array 1 is much similar to the complex pattern in the present study. It should be noted that the percentage of fixation on complex pattern of dots was the highest, though it was still under chance level.

Related studies [16] [20] indicated that pupil size was relevant to the mental workload of observers. Although there was a trend that average pupil size in the simple pattern and medium pattern conditions were larger than that in the complex pattern condition, the differences among the three pattern complexities were not significant. One possible explanation was that pupil size was not sensitive to the mental workload of observers, consistent with some prior study [16]. An alternative explanation was that the differences of mental workload under three pattern complexities were not strong enough.

Acknowledgement. This research was supported in part by grants from 973 Program of Chinese Ministry of Science and Technology (#2002CB312103) and the National Natural Science Foundation of China (#60433030 and #30500157).

References

1. Brockmole, J.R., Irwin, D.E.: Eye movements and the integration of visual memory and visual perception. Perception & Psychophysics 67, 495–512 (2005)
2. Brockmole, J.R., Wang, R.F.: Integrating visual images and visual percepts across time and space. Visual Cognition 10, 853–873 (2003)

3. Brockmole, J.R., Irwin, D.E., Wang, R.F.: The locus of spatial attention during the temporal integration of visual memories and visual percepts. Psychonomic Bulletin & Review 10, 510–515 (2003)
4. Brockmole, J.R., Wang, R.F., Irwin, D.E.: Temporal integration between visual images and visual percepts. Journal of Experimental Psychology: Human Perception & Performance 28, 315–334 (2002)
5. Di Lollo, V.: Temporal integration in visual memory. Journal of Experimental Psychology: General 109, 75–97 (1980)
6. Di Lollo, V., Hogben, J.H.: Suppression of visible persistence as a function of spatial separation between inducing stimuli. Perception & Psychophysics 41, 345–354 (1987)
7. Dixon, P., Di Lollo, V.: Beyond visible persistence: An alternative account of temporal integration and segregation in visual processing. Cognitive Psychology 26, 33–63 (1994)
8. Henderson, J.M., Weeks, P.A., Hollingworth, A.: The effects of semantic consistency on eye movements during complex scene viewing. Journal of Experimental Psychology: Human Perception and Performance 25, 210–228 (1999)
9. Hogben, J.H., Di Lollo, V.: Perceptual integration and perceptual segregation of brief visual stimuli. Vision Research 14, 1059–1069 (1974)
10. Hollingworth, A., Hyun, J., Zhang, W.: The role of visual short-term memory in empty cell localization. Perception & Psychophysics 67, 1332–1344 (2005)
11. Ichikawa, S.: Quantitative and structural factors in the judgment of pattern complexity. Perception & Psychophysics 38, 101–109 (1985)
12. Irwin, D.E.: Information integration across saccadic eye movements. Cognitive Psychology 23, 420–456 (1991)
13. Jiang, Y., Kumar, A., Vickery, T.J.: Integrating sequential arrays in visual-short term memory. Experimental Psychology 52, 39–46 (2005)
14. Jiang, Y., Olson, I.R., Chun, M.M.: Organization of visual short-term memory. Journal of Experimental Psychology: Learning, Memory, & Cognition 26, 683–702 (2000)
15. Just, M.A., Carpenter, P.A.: A theory of reading: From eye fixations to comprehension. Psychological Review 87, 329–354 (1980)
16. Lin, Y., Zhang, W.J., Watson, L.G.: Using eye movement parameters for evaluating human-machine interface frameworks under normal control opperation and fault detection situations. International Journal of Human-Computer Studies 59, 837–873 (2003)
17. Luck, S.J., Vogel, E.: The capacity of visual working memory for features and conjunctions. Nature 309, 279–281 (1997)
18. Pashler, H.: Familiarity and visual change detection. Perception & Psychophysics 44, 369–378 (1988)
19. Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychological Bulletin 124, 372–422 (1998)
20. Recarte, M.A., Nunes, L.M.: Mental workload while driving: effects on visual searclh, discrimination, and decision making. Journal of Experimental Psychology: Applied 9, 119–137 (2003)
21. Watson, D.G., Humphreys, G.W.: Visual marking: prioritizing selection for new objects by top-down attentional inhibition of old objects. Psychological Review 104, 90–122 (1997)
22. Watson, D.G., Humphreys, G.W.: Visual marking and visual change. Journal of Experimental Psychology: Human Perception and Performance 28, 379–395 (2002)

Event-Related Potential as a Measure of Effects of Report Order and Compatibility on Identification on Multidimensional Stimulus

I-Hsuan Shen[1], Kong-King Shieh[2], and Shin-Yuan Tsai[1]

[1] Department of Occupational Therapy, Chang Gung University, 259, Wen-Hwa 1st Road, Kwei-Shan, Tao-Yuan 330, Taiwan
{I-Hsuan Shen, shenih}@mail.cgu.edu.tw
[2] Department of Industrial Management, National Taiwan University of Science and Technology

Abstract. An experiment was conducted to evaluate the effect of order of reporting stimulus dimension in multidimensional stimulus identification using switch task paradigm. Nine healthy participants were required to identify each two-dimensional symbol by pushing the corresponding buttons on the keypad. The two orders of report were Order Color/Shape and Order Shape/Color. There was a task cue prior to each presentation of a symbol indicating the particular report order the participants should perform. The cue changed randomly. Results showed that order of report had a significant effect on response time for the first and second dimension. Analysis of behavioral data showed switch cost indicated by increase in response time was greater for Order Shape/Color, a less appropriate order of reporting dimensional values, than Order Color/Shape. It seemed plausible that participants needed more attention resource and showed more N2 inhibition for Order Shape/Color than for Order Color/Shape which fits the Chinese adjective-then-noun language habit.

Keywords: report order, multidimensional stimulus identification, task switch, event-related potential.

1 Introduction

The problems in designing visual displays have been among the most important topics in human factors engineering. Compacting information into a single multidimensional stimulus can be an effective way of utilizing limited display space and reducing clutter [1]. Operators search for dimensions in a particular order in a multidimensional stimulus-identification task implies that search performance is based on the specific order in which dimensions are examined [2].

1.1 Report Order

The order of reporting dimensional values may play an important role in the performance of identifying targets. Recently, Shieh and his colleagues [3] [4] investigated the effects of order of report on the speed and accuracy of identifying

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 591 – 599, 2007.

multidimensional stimuli. They found that subjects responded faster and more accurately if there was a natural language-appropriate order of reporting the dimensional attributes. The stimuli they used were a subset of the Naval Tactical Display System symbols [5]. Two dimensions were attached to each symbol. The first dimension was Shape. A symbol could be the circular, square, or angular. The second dimension was Part. A symbol could be upper half, full, or lower half. They found that the order of reporting dimensional values affected the speed and accuracy of identification. Subjects responded faster and more accurately if the order of reporting stimulus-dimension values was appropriate. The appropriate order of report is the order in which long-standing habits based on a standard word order are not violated. Reporting the Part dimension first then the Shape dimension is more consistent with the "adjective then noun" habit of native speakers of American or Chinese. For example, reporting a whole circular shape by "full circular" is more natural than reporting it by "circular full," hence giving effects of order of report.

1.2 Compatibility

The order of report in a multidimensional situation might be considered to be one type of S-R compatibility. Reporting the Part dimension first then the Shape dimension is more consistent with the "adjective then noun" habit of native speakers of American or Chinese. The order of report effect is a case of S-R compatibility; that is S-R compatibility is greater for Order Part/Shape than Order Shape/Part.

1.3 Task Switch

In task-switching paradigm, subjects are given two tasks. On some trials, subjects switch between the tasks while on others they repeat the previous task. A robust finding indicates that, when a task is performed on a switch trial (Task B then Task A), there is a sizable decrement in performance measured as increased reaction time or decreased accuracy compared with performance of a task which is repeated (Task A then Task A). This decrement is called switch cost and measured by reaction time [6] [7]. Gilbert and Shallice [8] suggested that the asymmetry of task switch cost is obtained when the two tasks demand larger differences in top-down control input, as in the classic Stroop tasks. It is easier to switch to the weaker task. The reverse pattern obtained when the tasks differ principally in the strength of their S-R mapping, as with typical S-R compatibility effects. If an experimental paradigm that amalgamates a task-switch design with two report orders as the two tasks is implemented. The appropriate order of report is the order in which long-standing habits based on a standard word order are not violated. The order of report effect is a case of S-R compatibility; that is S-R compatibility is greater for the appropriate order than the other. Based upon the results of the task-switching paradigm, it was assumed that switching from the report order of low S-R compatibility to the report order of high S-R compatibility is easier than switching in the opposite direction. In other words, if the report order of high S-R compatibility is indeed more appropriate, the effectiveness should be obtained by the

asymmetry of task switch cost. Switch cost will be greater for the report order of low S-R compatibility than the report order of high S-R compatibility.

1.4 Event-Related Brain Potential (ERP)

The event-related brain potential (ERP) has been used to provide a direct estimate of the timing of processes up to the intermediate stage of stimulus categorization [9] [10]. The ERP is a series of voltage oscillations or components recorded from the scalp to indicate the brain’s electrical response to discrete stimulus events. Stimulus features as well as the attention demands posed by a visual discrimination task may affect the amplitude, voltage topography across the scalp, and latency of ERP components [11]. The present study investigates the relationships between reaction times and ERP components for various report orders.

1.5 Hypotheses

The shape and the part dimensions used in the previous study [3] [4] do not appear to correspond to neurally based dimensions, which may have confounded the result. Attention research by Treisman [12] clearly showed that some dimensions are preattentively discriminated and have a special status (like color or orientation) whereas others are not. This builds on neural studies showing that dimensions like color and shape are processed in different parts of the brain. The present study aimed to investigate the effect of report order in multidimensional stimulus which was coded with color and shape dimension using a within-subjects design. It was hypothesized that response times would be longer for Order Shape/Color, switch trials, and switch cost would be greater from Order Color/Shape to Order Shape/Color than in the opposite direction.

2 Method

2.1 Participants

Nine male healthy college students (*M*=23 yr.) were tested. All had 18/20 corrected visual acuity or better and normal color vision. They were paid for their participation.

2.2 Stimuli

Nine symbols were used as the stimulus set (see Fig.1). Each of these nine symbols was encoded with two basic dimensions. The first dimension, "color," was red, yellow, or green. The second dimension, "shape," was circular, square, or triangular. The rim of each stimulus was colored with red, yellow, or green. Nine stimuli were used, presented against a black background. The height and width of the stimuli were about 1 cm by 1 cm.

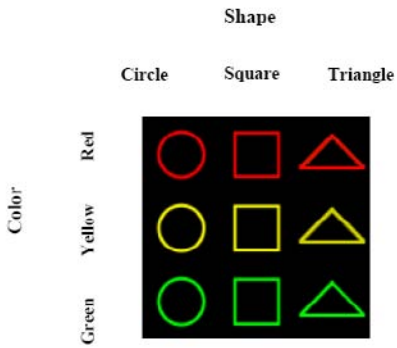


Fig. 1. One Nine symbols defined by color and shape dimensions. Each dimension had three values.

2.3 Procedure

The symbols were presented one per trial at the center of the display during the identification task. Viewing distance between the participant and display was approximately 60 cm. Before each trial, a participant fixated on a small cross on the middle of the screen. There was a task cue prior to each presentation of a symbol indicating which task the participant was to perform. For example, if the task cue was 'color/shape', then the participant had to report the 'color' dimension first and the 'shape' dimension second by pressing the appropriate buttons on the keypad. If the task cue was 'shape/color,' then the order of report was the reverse. The task cue changed randomly. On a non-switch trial, the task cue was preceded by the same task cue; on a switch trial, the task cue was preceded by a different task cue. The task cue was presented for 0.5 sec prior to the onset of each stimulus. Subjects completed four blocks of 72 trials, resulting in 288 trials for the entire experiment. A 2-min. rest occurred between blocks. The task cue was randomly selected and balanced for each stimulus. There were one practice blocks prior to the experiment. Participants were instructed to identify the symbol presented during each trial by pressing the buttons on the keypad which defined the symbols. Two columns on the keypad, three buttons in each column, were labeled with the descriptive names. The two columns represented the two stimulus dimensions, and the three buttons in each column represented the three values of that dimension. Each participant responded to the two stimulus dimensions by using the right index finger.

2.4 Design

The independent variables in the analysis were 2×2 within-subjects factors and included order of report (Order Color/Shape or Order Shape/Color) and trial type (switch trial or non-switch trial).

2.5 Electrophysiological Methods

The electroencephalogram (EEG) was recorded from 32 scalp locations using a Neuroscan Qcap32 AgCl electrode cap. The horizontal electrooculogram (EOG) was

recorded from the outer canthus of both eyes, and vertical EOG was recorded with electrodes located above and below the right eye. The EOG recordings were made to allow removal of eye-movement artifacts during data analysis. All recording sites were referred to linked mastoids. Interelectrode impedance was kept below 10 KOhm. EEG and EOG recordings were amplified using Neuroscan amplifiers at a bandpass of 0.1－70 Hz, and digitized at 500 Hz. A notch filter was used to remove 60 Hz interference.

2.6 Analysis of Data

Behavioral and ERP data were collected for the two conditions of Order Color/Shape and Order Shape/Color and for the two trial types of the switch and non-switch trial. Four behavioral measures were collected. Response time for the first stimulus dimension (RT_1) was the time between the presentation of a symbol and the participants' correct response to the first dimension. Response time for the second stimulus dimension (RT_2) was the time between participants' identification of the first dimension and their correct response to the second dimension. Response time total (RT_T) was the sum of RT_1 and RT_2. Total percentage correct was 100 times the number of symbols correctly identified divided by the number of symbols presented under each experimental treatment. The behavior data were submitted to a 2 × 2 analysis of variance with independent variables of report order (Order Color/Shape vs. Order Shape/Color) and trial type (switch vs. non-switch). The sum of square for subjects was subtracted from the error term. Analysis of variance was conducted separately for the four dependent measures.

ERP data on all trials were scanned offline for artifact contamination. Trials contaminated by eye blinks, muscle potentials, and other artifacts at any electrode, as well as incorrect responses were excluded from analysis. An artifact rejection criterion of ±50 μV was applied at all sites. Each epoch of EEG was from 200 msec. prestimulus (baseline) to 1000 msec. poststimulus. N1 was defined as a major negative wave that occurred between 50－130 msec.. N2was the major negative deflection between 240－400 msec. after stimulus onset. Mean amplitudes were measured in the two time windows. For ERP analysis, nine electrodes (F3, FZ, F4, C3, CZ, C4, P3, PZ, and P4) in the frontal, central, and parietal areas were selected. The N1 and N2 amplitude measures were subjected to three way analysis of variance (ANOVA) using the factors of report order (2 levels: Order Color/ Shape and Order Shape/Color), trial type (2 levels: switch and non-switch), and site (3 levels: frontal, central, and parietal).

3 Results

3.1 Behavior Data

Tables 1 to 4 summarize the means and standard deviations for the four dependent measures for the two orders of report and the two trial types of switch and non-switch trial separately. The switch cost was calculated by subtracting the mean reaction time on non-switch trials from the corresponding values on switch trials. The over-all mean of response time for the first dimension was 1160 msec. Analysis of variance showed

that order of report had significant effect on response time for the first dimension (RT_1). Response time for the first dimension for Order Color/Shape (1199 msec.) was significantly longer ($F_{1, 24}$=12.18, p < .05) than Order Shape/Color (1120 msec.). Though response time for the first dimension was greater for switch trial (1179 msec.) than for non-switch trial (1141 msec.), trial type showed no significant effect. Interaction between order of report and trial type was no significant. The overall mean of response time for the second dimension (RT_2) was 486 msec., much shorter than for the first dimension. Response times for the second dimension were 464 msec. and 509 msec. for Order Color/Shape and Order Shape/Color. The difference was statistically significant ($F_{1, 24}$=6.30, p < .05). Analysis of variance showed that the effects of trial type and interaction between order of report and trial type were not significant. The overall mean total response time (RT_T) was 1646 msec. Order of report showed no significant effect on total response time; however, the effect of trial type was statistically significant (switch trial: M = 1,678 msec.; non-switch trial: M = 1,614 msec.; $F_{1, 24}$=6.76, p < .05). The interaction between the two factors was not significant. Analysis of variance performed on the correct rate showed no significant effect for order of report (Order Color/Shape: 76.6%; Order Shape/Color: 75.1%).

Table 1. Means and Standard Deviations (msec.) of RT_1 for Each Experimental Condition

	Switch		**Non-Switch**		**Switch Cost**		
	M	*SD*	*M*	*SD*		*M*	*SD*
OrderColor/Shape	1209	285	1189	347	(20)	1199	308
Order Shape/Color	1148	305	1092	275	(56)	1120	283
	1179	288	1141	308	(38)	1160	294

Table 2. Means and Standard Deviations (msec.) of RT_2 for Each Experimental Condition

	Switch		**Non-Switch**		**Switch Cost**		
	M	*SD*	*M*	*SD*		*M*	*SD*
OrderColor/Shape	469	117	459	105	(10)	464	108
Order Shape/Color	528	182	489	137	(39)	509	157
	499	151	474	119	(25)	486	135

Table 3. Means and Standard Deviations (msec.) of RTT for Each Experimental Condition

	Switch		**Non-Switch**		**Switch Cost**		
	M	*SD*	*M*	*SD*		*M*	*SD*
OrderColor/Shape	1678	380	1648	422	(30)	1663	390
Order Shape/Color	1677	467	1581	374	(96)	1629	413
	1678	413	1614	388	(64)	1646	396

Table 4. Means and Standard Deviations of Correct Rate for Each Experimental Condition

	Switch		**Non-Switch**			
	M	*SD*	*M*	*SD*	*M*	*SD*
Order Color/Shape	75.9	5.1	77.3	5.9	76.6	5.4
Order Shape/Color	76.1	6.0	74.2	4.8	75.1	5.4
	76.0	5.4	75.7	5.5	75.9	5.4

3.2 Electrophysiological (ERP) Data

The ERPs (N1 and N4) at each of the nine sites (F3, FZ, F4, C3, CZ, C4, P3, PZ, and P4) were subjected to analysis of variance. Fig. 2 and 3 show the grand average of ERPs for Order Part/Shape and Order Shape/Part and for switch and non-switch trial, at FZ, CZ, and PZ. For the N1 component, the mean amplitudes for orders of report showed significant effects ($F_{1, 19}$=16.35, p < .05); however, the mean amplitudes for trial type showed no significant effects. N1 Mean amplitudes were more negative-going for Order Shape/Color. The N2 mean amplitudes showed significant effects for report order ($F_{1, 19}$=7.32, p < .005)). Mean amplitudes for N2 were more negative-going for Order Shape/Color than Order Color/Shape. There were marginal significant differences for trial type ($F_{1, 19}$=3.09, p =.07). Mean amplitudes for N2 were more negative-going for switch trials than non-switch trials.

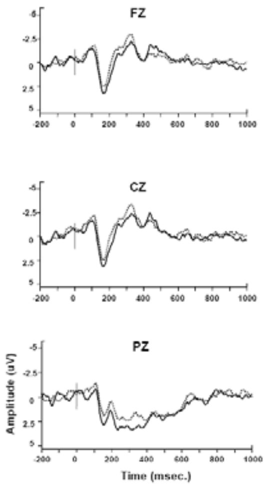


Fig. 2. Event-related potential waveform recorded for Order Color/Shape (solid line) and Order Shape/Color (dotted line) at FZ, CZ, and PZ electrode sites. Vertical line indicates symbol presentation.

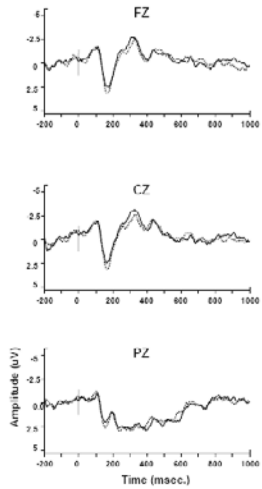


Fig. 3. Event-related potential waveforms recorded for switch trial (solid line) and non-switch trial (dotted line) at FZ, CZ, and PZ electrode sites. Vertical line indicates symbol presentation.

4 Discussion

This study was designed to examine the physiological evidences of the effects of report order using task-switch paradigm. Shieh, *et al.* [3] suggested that subjects responded faster and more accurately if the order of reporting stimulus dimension was natural language-appropriate. The suggestion only stood for RT_2 in this experiment but not for RT_1 and $RT_{T.}$ The present results showed that participants responded faster for the first stimulus dimension (RT_1) if they reported shape dimension first and color dimension second (Order Shape/Color) than if they report dimensions in the opposite order (Order Color/Shape). The result seemed contrary to our hypothesis. Only the rim of each

stimulus was colored and the size of each stimulus was very small, thus the color dimension was more difficult to be discriminated than shape dimension for participants. If each stimulus was colored in whole stimulus shape, the effect of report order might be the same as the studies of Shieh, *et al*. [3] for the first stimulus dimension (RT_1). With respect to the effect of task switch, it showed that RT_T was significantly greater for switch trials than for non-switch trials. Switch cost for RT_1 and RT_T was greater from Order Color/Shape to Order Shape/Color than in the opposite direction. Our hypothesis that the switch cost would be greater for Order Shape/Color than for Order Color/Shape was supported by these data. The results indicated that the Order Color/Shape is like natural appropriate language habits and so of greater S-R compatibility.

The N1 has been described as a nonunitary process, reflecting different psychophysiological functions, including the state of subject's arousal [13] and being augmented during attention to visual stimulation. Picton, *et al*. [14] suggested that N1 might represent the processes necessary for the evaluation of incoming information. Vogel, *et al*. [15] found a larger N1 wave for choice-RT tasks, and this effect is equivalent for form and color discrimination, indicating that the process reflected by the N1 discrimination effect is more general. In the present study, N1 mean amplitudes were significantly more negative-going for Order Shape/Color than for Order Color/Shape. This might reflect participants were more attentive for Order Shape/Color. The N2 component was the major negative deflection between 240 and 400 msec. after stimulus onset. By using the classical go/nogo task several ERP researchers claim to have found ERP correlates of a frontal inhibition mechanism [16] [17]. Compared with the ERPs elicited by go trials, nogo trials show a greater frontocentral negative wave peaking around 200-400 msec. (nogo-N2) [16] [17]. That is, in a nogo trial, subjects need to inhibit the tendency to respond and the inhibition results in a greater nogo-N2. Report order effect was found for N2 amplitude. Result showed that mean amplitude of N2 was greater negative magnitude in Order Shape/Color than Order Color/Shape. The N2 amplitude was maximal at the frontocentral electrode sites, similar to the findings of previous go/nogo studies. The less appropriate Order Shape/Color induced greater N2 amplitude. Order Color/Shape is more natural and of greater S-R compatibility than Order Shape/Color. When participants responded in lesser compatible Order Shape/Color, they showed greater N2 inhibition because they had to inhibit the tendency to respond in the more natural and compatible Order Color/Shape.

This research used behavioral and physiological indices to examine the effects of order of report and task switch on multidimensional stimulus identification. The task switch cost for Order Shape/Color was greater than that for Order Color/Shape. Results indicated that Order Shape/Color might be less compatible than Order Color/Shape. Physiological data showed that Order Shape/Color seemed to elicit greater attentive demand and inhibition.

References

1. Tsang, P.S., Bates, W.E.: Resource of allocation and object displays. In: Proceedings of the Human Factors Society 34th Annual Meeting, pp. 1484–1488. Human Factors Society, Santa Monica, CA (1990)

2. Fisher, D.L., Tanner, N.S.: Optimal symbol set selection: a semiautomated procedure. Human Factors 34, 79–95 (1992)
3. Shieh, K-K., Lai, C-J., Ellingstad, V.S.: Effects of report order, identification method, and stimulus characteristics on multidimensional stimulus identification. Perceptual and Motor Skills 82, 99–111 (1996)
4. Shieh, K-K., Chen, F-F.: Effects of report order and stimulus type on multidimensional stimulus identification. Perceptual and Motor Skills 95, 783–794 (2002)
5. Osga, G.: An evaluation of identification performance for Raster Scan generated NTDS symbology (Final Report). Naval Ocean System Center, San Diego, CA (1982)
6. Allport, A., Styles, E.A., Hsieh, S.: Shifting Intentional Set: Exploring the Dynamic Control of Tasks. In: Umilta, C., Moscovitch, M. (eds.) Attention and Performance: 15 Conscious and Nonconscious Information Processing, pp. 421–452. MIT Press, Cambridge, MA (1994)
7. Rogers, R.D., Monsell, S.: Costs of a predictable switch between simple cognitive tasks. Journal of Experimental Psychology: General 124, 207–231 (1995)
8. Gilbert, S.J., Shallice, T.: Task switching: a PDP model. Cognitive Psychology 44, 297–337 (2002)
9. Donchin, E.: Surprise! surprise! Psychophysiology, Vol. 18, 493–513 (1981)
10. Gratton, G., Coles, M.G.H., Sirevaag, E., Eriksen, C.W., Donchin, E.: Pre- and post-stimulus activation of response channels: a psychophysiological analysis. Journal of Experimental Psychology: Human Perception and Performance 14, 331–344 (1988)
11. Hillyard, S.A., Mangun, G.R., Woldorff, M.G., Luck, S.J.: Neural system mediating selective attention. In: Gazzaniga, M.S. (ed.) The cognitive neurosciences, pp. 665–681. MIT Press, Cambridge, MA (1995)
12. Treisman, A.: Feature binding, attention and object perception. Philosophical Transactions of the Royal Society, Series B 353, 1295–1306 (1998)
13. Näätänen, R., Picton, T.: The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24, 375–425 (1987)
14. Picton, T.W., Campbell, K.B., Baribeau-Braun, J., Proulx, G.B.: The neurophysiology of human attention: a tutorial review. In: Requin, J. (ed.) Attention and performance VII, pp. 429–467. Lawrence Erlbaum, Hillsdale, NJ (1978)
15. Vogel, E.K., Luck, S.J.: The visual N1 component as an index of a discrimination process. Psychophysiology 37, 190–203 (2000)
16. Pfefferbaum, A., Ford, J.M., Weller, B.J., Kopell, B.S.: ERPs to response production and inhibition. Electroencephalography and Clinical Neurophysiology 60, 23–34 (1985)
17. Kok, A.: Effects of degradation of visual stimuli on components of the event-related potentials in Go/Nogo reaction tasks. Biological Psychology 23, 21–38 (1986)

Models of Command and Control

Neville A. Stanton, Guy Walker, Dan Jenkins,
Paul Salmon, Mark Young, and Amerdeep Aujla

Defence Technology Centre for Human Factors Integration,
BIT Lab, School of Engineering and Design, Brunel University,
Uxbridge, UB8 3PH, UK

Abstract. This paper reports on four different models which may be used to investigate command and control. The models reviewed are: a process model, a contextual control model, a decision ladder model and a functional model. Each of the models are introduced and explained in turn. Of particular interest is the degree to which these models can be used to explain command and control activities. The models are compared and contrasted, and the implications for command and control research drawn.

Keywords: Command and control, process models, contextual models, decision models, functional models, activities models.

1 Introduction

The purpose of this paper is to investigate alternative models of command and control. In particular, the paper presents the development of an activity model and contrasts this with control, contextual, decision and functional models that have been proposed by other researchers. The purpose of the activity model is to provide a research tool that may be applied to any command and control domain. It is claimed that this model will enable researchers to conduct investigations into different aspects of command and control in a systematic manner. The main aims of any model are to simplify complexity and to provide the basis for predictions of performance of the phenomena under consideration.

Four contemporary models of command and control are presented as a way of summing up the field. The first is a control theoretic model by Lawson (1981), the second is a control modes model by Hollnagel (1993), the third is the decision ladder model by Rasmussen (1974) and Vicente (1999), and the fourth is a functional command and control model by Smalley (2003). As there is approximately a decade between the publications of these models, they are probably representative of the changes in thinking over the past three decades, although each model only represents a snapshot of each particular approach.

2 Process Model

Command and control may be viewed as an information processing chain, as data flows between the environment, one's own forces and the command centre. The

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 600–608, 2007.

model in Figure One epitomises this perspective. The model is rooted in the idea that there is some desired state that the command centre seeks to achieve. Data are extracted from the environment and processed. The understanding of these data are then compared with the desired state. If there is any discrepancy between the desired state and the current state, the command centre has to make decisions about how to bring about the desired state. These decisions are turned into a set of actions, which are then communicated to their own forces. The data extraction cycle then begins afresh.

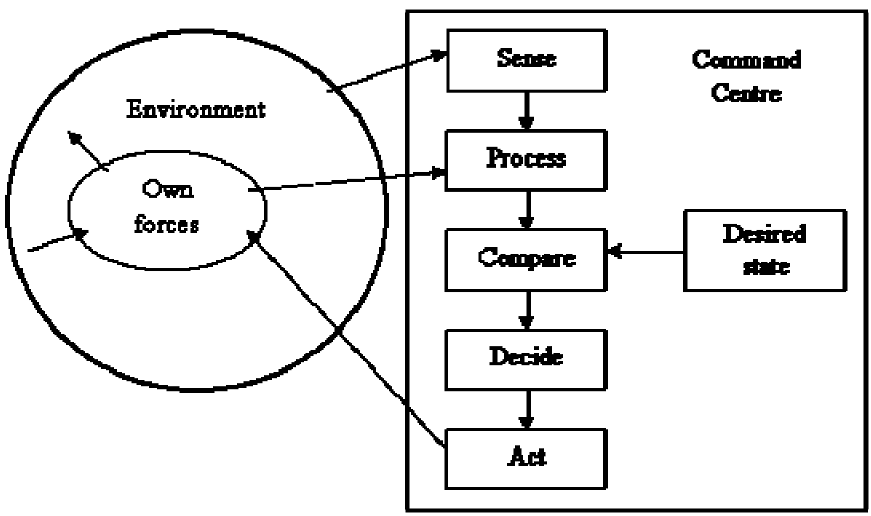


Fig. 1. An adapted version of Lawson's model of the command and control process. From: Lawson, J. S. (1981) Command and control as a process. IEEE Control Systems Magazine, March, 86-93.

Lawson's model owes much to the ideas from control theory. The comparison of actual and desired states implies a feedback process and some form of regulation. Central to his model, therefore, would be the "compare" function. The feedback involves control of "own forces" to affect a change to the environment. The notional "actual" and "desired" states imply phenomena that can be described in terms of quantitative, discrete data; in other words it is not easy to see how the model would cope if the actual state was highly uncertain. Nor is it easy to see what would happen if the changes to the environment led to consequences which lay outside the limits defined by the discrete state. The model does indicate the central issue that command can be thought of as working towards some specified effect or intent but suffers, however, from its apparent reliance on a deterministic sequence of activities in response to discrete events.

3 Contextual Model

Hollnagel (1993) developed a Contextual Control approach to human behaviour, based on cognitive modes, to explain the effects of the context in which people performed their actions. Rather than command and control being a pre-determined sequence of events, Hollnagel has argued that it is a constructive operation where the operator actively decides which action to take according to the context of the situation together with his/her own level of competence. Although set patterns of behaviour maybe observed, Hollnagel points out that this is reflective of both the environment as well as the cognitive goal of the person, both of which contain variability. In the Contextual Control Model, shown in figure two, four proposed modes of control are as follows.

- Strategic Control - is defined as the ‘global view’, where the operator concentrates on long term planning and higher level goals.
- Tactical Control - is more characteristic of a pre-planned action, where the operator will use known rules and procedures to plan and carry out short term actions.
- Opportunistic Control - is characterised by a chance action taken due to time, constraints and again lack of knowledge or expertise and an abnormal environmental state.
- Scrambled Control - is characterised by a completely unpredictable situation where the operator has no control and has to act in an unplanned manner, as a matter of urgency.

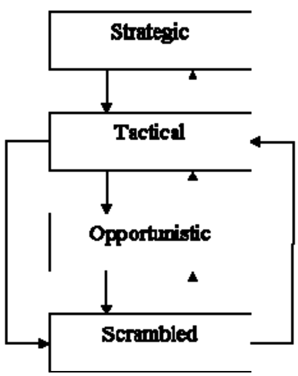


Fig. 2. Hollnagel’s contextual control model. From: Hollnagel, E. (1993). Human reliability analysis: Context and control. London: Academic Press.

The degree of control is, therefore, determined by a number of varying interdependent factors. Hollnagel considers that availability of subjective time is a main function of command and control - this means that as the operator perceives

more time available so s/he gains more control of the task/situation. The factors affecting the perception of available time may include: the number of goals, the availability of plans to meet these goals, the modes of execution of those plans and the time available. At any point in time the system operator is attempting to optimise all of these criteria. Hollnagel's model differs from Lawson's, in that it does not prescribe the sequence and relations of command and control activities; rather it proposes contextual differences in the control mode. There is some evidence to support this hypothesis. In a study of team behaviour in a supervisory control task, Stanton et al (2001) showed that the transitions between control modes were consistent with Hollnagel's model.

4 Decision Model

The Decision Ladder model of activity was developed by Rasmussen (1974) who observed that expert users were relying on a mixture of knowledge-based, rule-based and skill-based behaviour to conduct tasks. Rasmussen proposed that the sequence of steps between the initiating cue and the final manipulation of the system can be identified as the steps a novice must necessarily take to carry out the sub task, as shown in figure three. Studies of expert performance may then result in a description of his performance in terms of shunting leaps with this basic sequence. The ladder can be seen to contain two different types of node: information processing activities (represented by the rectangular boxes) and knowledge states (represented by the circles). By folding this list in half it is possible to add links between the two sides. There are two types of shortcut that can be applied to the ladder; 'shunts' connect an information-processing activity to a state of knowledge (box to circle) and 'leaps' connect two states of knowledge (circle to circle). This is where one state of knowledge can be directly related to another without any further information processing. It is not possible to link straight from one box to another, as this misses out the resultant knowledge state (Vicente, 1999).

According to Naikar & Pierce (2003) the left side of the decision ladder represents an actor observing the current system state whereas the right side of the decision ladder represents an agent planning and executing tasks and procedures to achieve a target system state. Sometimes observing information and diagnosing the current system state immediately signals a procedure to execute. This means that rule based shortcuts can be shown in the centre of the ladder. On the other hand the actor may have to engage in effortful, knowledge-based goal evaluation to determine the procedure to execute; this is represented in the top of the ladder. There are two types of shortcut that can be applied to the ladder; 'shunts' connect an information-processing activity to a state of knowledge (box to circle) and 'leaps' connect two states of knowledge (circle to circle). The path in which the operator moves through the ladder is dependant on a number of factors including; their workload, experience and familiarity with the current task.

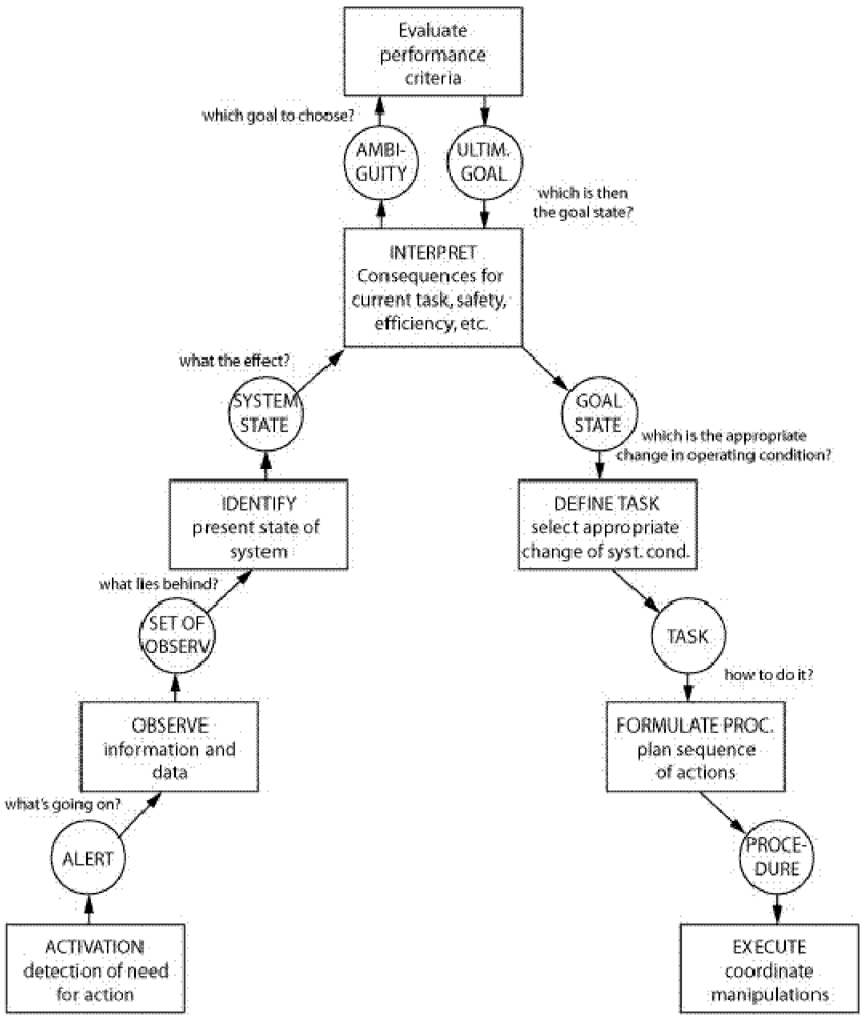


Fig. 3. Rasmussen and Vicente's decision ladder model. From: Rasmussen, J. (1974). The human data processor as a system component: Bits and pieces of a model (Report No. Risø-M-1722). Roskilde, Denmark: Danish Atomic Energy Commission.

5 Functional Model

Smalley (2003) proposed a functional model of command and control, comprising some seven operational and decision support functions (six in the ovals and one in the box) and ten information processing activities (appended to the input and output arrows). The ten information processing activities are: primary situation awareness, planning, information exchange, tactical situation reports, current situation awareness,

directing plan of execution, system operation, system monitoring, system status, and internal co-ordination and communications. A representation of the relationship between the operation and decision support functions and information processing activities is shown in figure four.

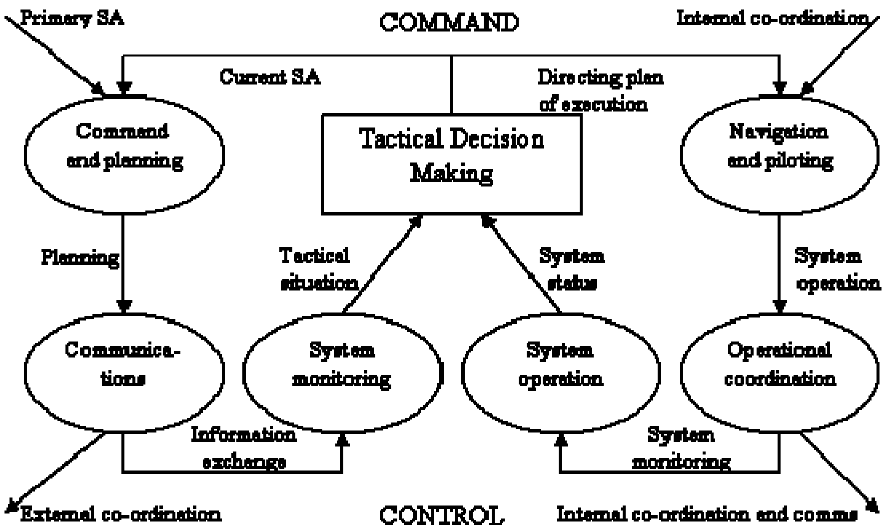


Fig. 4. Smalley's functional command and control model. From: Smalley, J. (2003) Cognitive factors in the analysis, design and assessment of command and control systems. In: E. Hollnagel (ed) Handbook of Cognitive Task Design. Mahwah, New Jersey: Lawrence Erlbaum Associates (223-253).

Information about the state of the world is collected through the primary situation awareness activities. The various sources of information are combined so that targets and routes can be defined in the planning activities. Information about targets, routes and intentions is exchanged with other forces. The status of the mission is communicated through the tactical situation reports. Current situation awareness activities merge information about the mission with primary situation awareness, to inform the planning process. The information from this latter set of activities will cue the start of the activities that direct the plan of execution. This, in turn, informs activities associated with the direction of system operation. The system is monitored, to see if outcomes are as expected. Any changes in the system status may lead to changes in the planning and the directing of the plan. Internal and external co-ordination and communication activities keep the command and control system functioning.

Smalley's model comprises an integration of many command and control activities with feed-forward and feed-back loops. It has a higher level of command and control fidelity than the other two models. The model suggests that 'command' activities (at the top of the figure) are separate, but connected to, the 'control' activities (at the bottom of the figure). The activities on the right-hand side of the figure are concerned

with internal operation of the system, whereas the activities on the left-hand side of the figure are concerned with interfacing with the external environment. In contrast to the other three models, Smalley's model offers high much higher fidelity for command and control.

6 Conclusions

Comparison of the models of command and control reveals some differences in those models considered earlier. The four models presented were Lawson's (1981) control theoretic model, Hollnagel's (1993) control modes model, Rasmussen (1974) and Vicente's (1999) decision ladder model, and Smalley's (2003) functional command and control model. The model in figure four contains all of the information processing activities within Lawson's model (shown in figure one), but with greater fidelity and relevance to command and control activities. There is no explicit representation of a 'desired state' in the newer model (see figure four); rather this is expressed in terms of 'required effects' which may be open to change in light of changes in the mission or events in the world. It is also worth noting that the 'desired state' appears to be static in Lawson's model, which is a potential weakness of the approach.

Whilst much of the command and control activities are implicitly listed in Hollnagel's model (shown in figure two) under 'goals', 'plans', 'execution' and 'events', the new model makes all of these activities explicit (see figure four). The model does not indicate the effects of temporal change on the command and control activities. The model does however distinguish between the proactive 'command' activities and the reactive 'control' activities. It is probable that in higher tempo situations the command and control system is more likely to be in a reactive mode (i.e., the right-hand side of figure four). Conversely, it is probable that in lower tempo situations the command and control system is more likely to be in a proactive mode of operation (i.e., the left-hand side of figure four). The newer functional model in figure four does not attempt to distinguish between knowledge and system states in the way that Rasmussen and Vicente's decision ladder model does. The decision ladder model has the inherent flexibility of shortcuts (which the analyst is required to identify), which can be used to indicate different levels of expertise. This makes the decision ladder model a better explanation of individual behaviour, but less applicable to the description of the activities in a command and control system. The generic decision ladder model could be used to describe any system, but it lacks the fidelity of Smalley's model for command and control. The decision ladder model does not attempt to distinguish between command and control activities, nor proactive and reactive behaviour. Smalley's model (shown in figure four), does distinguish between 'command' activities and 'control' activities, and between 'internal' and 'external' co-ordination. Smalley's model also has a high degree to fidelity with regard to the command and control activities. This was seen as a significant strength of the model. Whereas Smalley's model was based solely on military command and control the previous models were based on a broader domain base. All of the models provide a basis for research investigations into command and control activities. Given the

various models above, the relatively invariant properties of command and control scenarios can be distilled down to a basic descriptive level as a scenario possessing:

- a common overall goal (comprised of different but interacting sub-goals),
- individuals and teams coordinating to reach it, but
- dispersed geographically, and there are
- numerous systems, procedures and technology to support their endeavour (Walker et al., 2006).

Beyond the descriptive level, command and control by definition is a collection of functional parts that together form a functioning whole. Command and control is a mixture of people and technology, typically dispersed geographically. It is a purposeful intelligently adaptive endeavour representing progress towards a defined outcome. Intelligent adaptation requires responses to externally generated input events within a finite and specified period (Young, 1982). In possessing these attributes, command and control can be characterised with reference to, and understood from the following modelling perspectives, as:

- An (open) system of interacting parts,
- A socio technical system of human and non-human agents and artifacts,
- A distributed system,
- A real time system, and
- An ‘intelligent’ system.

There are numerous modelling challenges underlying command and control and the attributes above. Foremost is that the, “real world is made from open, interacting systems, behaving chaotically” (Hitchins, 2000), and in the case of human actors, non-linearly. Complex systems like command and control scenarios also possess various real-time properties that cannot be considered ‘designed’ as such, they sometimes merely ‘happen’ (Hitchins, 2000). Therefore the notion of a commander representing something akin to the conductor of an orchestra is in some cases entirely false (Hitchins, 2000). Also, unlike clearly linear systems, the possibility exists for there to be no clear boundaries between certain system elements, as well as no beginning and no end, given that goals are more or less externally adaptive. Within this, the concept of situational awareness arises as a “necessary component in achieving decision superiority” (p. 28), and one of the key emergent properties from any command and control scenario (DoD, 1999). Added to the previous challenges, in virtually anything other than the physical sciences a theory such as situational awareness is more akin to, “a general principle or a collection of interrelated general principles that is put forward as an explanation of a set of known facts and empirical findings” (Reber, 1995, p. 793). Thus the modelling of a generic command and control system as an entity is very challenging; as this is a complex social-technical environment that aims to promulgate a shared situational understanding across dispersed geographical locations, using various technical and communications media. By conducting observations across several domains, the aim of this work has been to develop a generic framework for command and control. In order to progress this into a coherent theory, the next phase of the work is to explore how the various domains perform operations within each heading and to ask how the removal or disruption of activity under a heading will impair performance within a given domain.

References

1. DoD: Realizing the potential of C4I: Fundamental Challenges. Washington, National Academy Press, DC (1999)
2. Hollnagel, E.: Human reliability analysis: Context and control. Academic Press, London (1993)
3. Hitchins, D.K.: (2000), http://www.hitchins.net/CandC.html
4. Lawson, J.S.: Command and control as a process. IEEE Control Systems Magazine, pp. 86–93 (March 1981)
5. Naikar, N., Pearce, B.: Analysing activity for future systems (2004) Defence Science and Technology Organisation (Unpublished report)
6. Rasmussen, J.: The human data processor as a system component: Bits and pieces of a model (Report No. Risø-M-1722). Danish Atomic Energy Commission, Roskilde, Denmark (1974)
7. Reber, A.S.: Dictionary of Psychology. Penguin, London (1995)
8. Smalley, J.: Cognitive factors in the analysis, design and assessment of command and control systems. In: Hollnagel, E. (ed.) Handbook of Cognitive Task Design, pp. 223–253. Lawrence Erlbaum Associates, Mahwah, New Jersey (2003)
9. Stanton, N.A., Connelly, V., Prichard, J., van Vugt, M.: Assessing the effects of location, media and task type on team performance. Journal of Defence Studies 7(1), 29–41 (2002)
10. Vicente, K.J. (ed.): Cognitive work analysis: Toward safe, productive, and healthy computer-based work. Lawrence Erlbaum Associates, Mahwah, NJ (1999)
11. Young, S.J.: Real time languages: Design and Development. Ellis Horwood, Chichester (1982)

Cognitive and Emotional Human Models Within a Multi-agent Framework

Lucas Stephane

European Institute of Cognitive Sciences and Engineering (EURISCO International)
4 Avenue Edouard Belin 31400 Toulouse, France
lucas.stephane@eurisco.org

Abstract. In the industrial field, user behavior has been mainly addressed in terms of rational thinking. High-level cognitive processes investigated by Cognitive Engineering are described as logical or rational. We already proposed a cognitive multi-agent model that provides a framework for the peer user-artifact highlighting the roles, responsibilities and resources of each pertinent entity involved in the human-machine hybrid system. This cognitive model was employed for various experiments in the cockpit that assessed Workload and Situation Awareness [10]. Techniques such as Eye Tracking were also used. Definitely, the cognitive model helps to understand user behaviors. However, we noticed behavioral differences between users that are hardly explainable only by the cognitive model. So we started to investigate the non-cognitive aspects of the users that are their emotions. This paper presents the integration of cognitive and emotional models that comprise on one hand users' Situation Awareness and on the other hand users' Self Awareness.

Keywords: Cognitive modeling, Cognitive theories of emotion, Multi-Agent systems, Situation Awareness, Workload, Self Awareness.

1 Introduction

Human knowledge continues to expand day after day in both natural and technical domains. As this knowledge grows, reliable formalisms able to deal with complexity are needed. A good support has been provided by the information technologies. In this field, such formalisms evolved from function based to object based and finally to multi-agent systems. Roughly, the first formalisms rely on data input, data process and data output, like elementary functions in mathematics. The second formalism define entities as objects, each object having a set of methods that operate on its' attributes. The third type of formalism provides a societal or agency structure, using entities called agents. Agents are defined by their roles, responsibilities, resources and goals [4]. This last formalism is suitable for large and complex systems.

We employed a multi-agent formalism for our cognitive engineering purposes, i.e. we specified cognitive agents, and we identified their roles, responsibilities, resources and goals. We used this multi-agent approach as a generic container with the aim to integrate the cognitive theories and methods that we use in the design-evaluation cycle. These methods are the Cognitive Function Analysis *(CFA)* [2], Skills, Rules

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 609–618, 2007.

Knowledge *(SRK)* [7], Cognitive Compatibility Situation Awareness Rating Technique *(CC-SART)* [11], Situation Awareness Global Assessment Technique *(SAGAT)* [3], NASA Total Load Index *(NASA TLX)* and Modified Cooper-Harper.

Working in the applied research field, our goal is to provide a framework for operational use. This framework may be considered as a modular set of tools that have to be used according to the industrial targets and goals. At the same time, this framework ensures completeness and traceability in space and time, aiming to integrate in the end results obtained possibly independently via each tool in the iterative design-evaluation loop.

The context of use has been studied mainly as external to the user (from physical comfort to Human Machine Interfaces), and Workload and Situation Awareness theories and methods used today have been set up related to this external environment. However, the user model should be investigated also in terms of introspection and internal states that define finally the users' ability to act, that we call here Self Awareness. Taking into account the need to consider emotional aspects of the user [9], we started to investigate theories of emotion that may be used in the industrial field and, as a continuum of an integrated approach, we tried to complete the user model using the same multi-agent formalism. The two emotion theories selected are the Schematic, Propositional, Analogical and Associative Representation Systems (*SPAARS*) [6] and the Self-Regulatory Executive Function (*SREF*) [12]. The first one addresses the relations between cognition and emotion, while the second one addresses the relations between attention and emotion.

We think both Situation Awareness and Self Awareness enable finally decision making and action.

2 Cognition Theories and Models

The rationale of using cognitive approaches in order to understand, design and assess Human Machine Interaction has been largely proved. However, even if these last times several sciences such as neurosciences, information technologies, psychology and philosophy are aggregated under the 'Cognitive' label, the main use of cognition in industry is based on rational thinking, logic, rules and procedures. In this paper, cognition will be used in these last terms. This section presents the set of cognitive theories that we currently use.

2.1 Skills, Rules and Knowledge and Cognitive Compatibility Situation Awareness Rating Technique

A very well known cognitive model is Rassmussen's Skills, Rules and Knowledge [7]. In this model, human behaviors and activities rely on three levels that go bottom-up from automatic behaviors to rule-based and finally to knowledge-based behaviors. The distribution of behaviors to these three levels may change in time, top-down, rules emerging from knowledge and automation emerging finally from rules.

The Cognitive-Compatibility Situation Awareness Rating Technique [11] adapts SRK model to explain and assess Situation Awareness. The three main dimensions of CC-SART are the Level of Processing, the Ease of Reasoning and the Activation of Knowledge. Their direct relation to SRK is illustrated in Table 1 below. These two theories are centered on the understanding of the situation.

Table 1. The three dimensions of CC-SART related to the SRK model

CC-SART	SRK
Level of Processing The degree to which the situation involves natural, automatic, intuitive and associated processing	**Skill-Based Behavior** Activities take place without conscious control as smooth, automated, and highly integrated patterns of behavior.
Ease of Reasoning The degree to which the situation is straightforward and understandable and not confusing and contradictory. It is based on the understanding of the rules or procedures provided to the user and on the degree of affordance inherent to the artifact and/or situation.	**Rule-Based Behavior** A familiar work situation is typically controlled by a stored rule or procedure that may have been derived empirically during previous occasions. The higher level rule-based co-ordination is generally based on explicit know-how, and the rules used can be reported by the person.
Activation of Knowledge The degree to which the situation is recognizable and familiar, or is strange and unusual. The activated knowledge may be related to the domain of use of the artifact, or may be cross-domain and associative, if the situation is new to the user.	**Knowledge-Based Behavior** During unfamiliar situations, faced with an environment for which no know-how or rules for control are available from previous encounters, the control of performance must move to a higher conceptual level, in which performance is goal-controlled and knowledge-based.

2.2 Situation Awareness Global Assessment Technique

In order to behave appropriately, the user has not only to perceive information that enable to understand the current situation, but also to anticipate the evolution of the situation related to his own actions on the artifact, as well as to the context influence or impact on the artifact (i.e. wind impact on the plane trajectory). Thus, in SAGAT [3], Situation Awareness is declined on three levels – perception, understanding and projection – that are linked to the active goal and dynamically regulated by the Mental Models of the user.

Mental Models [2], [3], [6], are defined as representations of past experiences used in a predictive way. A Mental Model is generic. A Mental Model is the representation of the artifact features and contextual events and states that enable the user to mentally try out actions before executing them. The operative structure of the artifact is built as an operative mental mode or image [2].

A schema is an instance of a Mental Model for a specific system and situation. Schema [3], [6] are packets of knowledge, tuned by experience.

2.3 Cognitive Function Analysis

Cognitive Function Analysis [2] is an integrated task analysis method that tightly ties the user, the task, the artifact and the organizational environment. The role of a cognitive function or of a set of cognitive functions is to transform a prescribed task into effective activity. A Cognitive Function has a role, a context and a set of

resources. Thus, Cognitive Functions have a teleological definition, i.e. the function of the eye is to see, rather than a mathematical definition. As such, Cognitive Functions are considered as agents. CFA provides powerful methods to describe and assess the context and the situation – context patterns and situation patterns; perceived versus desired situation –, the differences between novices and experts – the Situation Recognition and Analytical Reasoning model (*SRAR*) model –, as well as the Human Machine Interaction. For this purpose, Interaction Blocks are defined by triggering preconditions, action, context pattern, abnormal conditions and goal. CFA focus the automation issues, and provides solutions for the transfer and balancing of Cognitive Functions between the user and the system. Concerning the organizational environment, CFA proposes the Active Design Documents (*ADD*) that insure traceability through the various iterations of the Human Centered Design process. Active Design Documents are chunks of technical documentation defined by interaction descriptors, interface objects and contextual links. Each ADD contains four sections that are the Design Rationale, the Task, the Artifact (or prototype) and the Evaluation.

3 Emotion Theories

Since a long time, several psychological theories dealt with human emotions [8]. Meanwhile, as long as they have not been validated with scientific arguments, they have not been considered in the engineering field. But since the mid 90s when the neurosciences brought serious evidence for emotional circuits [5], there is a renewal of interest in investigating emotions and the way they impact on human activity. For the consumer products for example, a new domain called Kansei Engineering (Kansei means ‘total emotion’ in Japanese [9]) provides systematic methods to design and assess the emotional impact of the product on users and on their decision to acquire or use the given product. However, in safety-critical domains, we need to investigate emotions deeper, and especially how emotion and cognition interrelate. In this section we define emotions, expose their rationale in the industrial field, and finally we propose two theories of emotion that might be used to complete the user model defined only in cognitive terms.

3.1 Emotion Definition

The word emotion comes from Latin *motere* that means movement. Thus emotion is immediately linked to action, and each basic or fundamental emotion has a vital role (function). There are different limited definitions of emotion, depending on the field of investigation [8]. Affective states are considered as the superset of emotions, feelings and humors. As a *state*, an emotion has a precise start, is linked to a specific object and has a given duration. The emotional *process* is the sequence of several emotional states. Emotions are characterized by physiologic activations, and the awareness of her (his) own physiologic changes becomes a cognitive stimulus for the subject. According to Ekman [8], a fundamental emotion has a universal and distinctive signal, is common to other primates than the humans, has a specific template of physiologic reactions, is associated to universal and specific triggering

events, is fast detected, has a short life cycle, is appraised automatically and appears spontaneously. Panskepp [8] takes into account the neurophysiologic aspects of fundamental emotions. Thus, fundamental emotions rely on sensory-motor command circuits that are structured according to the circumstances and that are critical for the actions fundamental to survive. Located in the di-encephalic and in the limbic systems, these circuits coordinate behavioral reactions, autonomous and hormonal processes and subjective states or moods.

3.2 Rationale of Studying Emotions

Emotions may contribute to explain the differences between the prescribed task and the real activity. For the visual channel, emotion modeling could be directly linked to Situation Awareness. Emotion models contribute to better interpret physiological data such as heart rate that is currently an indicator of workload.

Together with a Swiss military pilot working at Eurisco, we identified several topics of the human factors domain where emotions might be taken into account:

- *Relationship User – System*: the emotional state of the subject depends on the degree of visibility and context understanding; emotions express also the degree of trust and security offered by the system, and are crucial in situations such as system failures and malfunctions
- *Graphical User Interface Design*: identification and implementation of invariable items, whatever the users' emotional state; assess user emotions when facing virtual versus real world scenes, i.e. simulators versus real situations; implement functions that appease the user
- *Training*: identification of emotions and comparison of experts and novices; investigate anxiety and risk correlated to expertise; implementation of risk management tools
- *Team collaboration*: identify the role of emotions in non-verbal communication, i.e. face expressions, gestures; analysis of co-worker expressed emotions as triggers for information gathering on the user interface; identification of non-verbal communication, i.e. gestures and facial expressions, intra-team patterns related to the context, i.e. extreme conditions versus normal conditions; comparison of information gathering on the co-worker(s) and on the system

The next step was to find theories of emotion that may be adapted to the industrial research, and furthermore integrated with the cognitive theories and methods already used. Two theories developed in the cognitive therapy field retained our interest. The first one focuses the relationships between cognition and emotion, and the second one focuses the relationships between attention and emotion.

3.3 Schematic, Propositional, Analogical and Associative Representation Systems *(SPAARS)*

SPAARS [6] uses four levels of representation. The first level is the analogical one that includes the sensory-specific systems: visual, auditory, tactile, proprioceptive and olfactory. The output of the analogical processing is sent to three semantic representation levels that work in parallel (Figure 1).

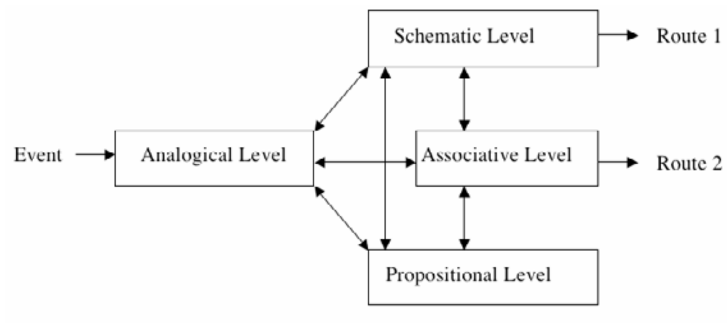


Fig. 1. The four representation levels of SPAARS

The lower one, fully automated, is associative (Route 2). Automatization of emotions within SPAARS occurs during cycles of appraisal at the highest level (hereafter) with the repetition of event-emotion combinations. Once automated, emotions are difficult to modify and happen invariantly in the presence of the triggering event. The intermediate level is propositional. This propositional language-like representation does not propose a direct route from proposition to emotion, but is linked either through appraisals to the highest schematic level (hereafter), either directly through the associative route, i.e. particular words directly linked to emotion such as swear words. The highest level is schematic and corresponds to the Mental Model representation. In relation to emotion, this level is very important because at this level occurs the generation of emotion through the process of interpretation and appraisal of any relevant input, of external or internal origin (Route 1).

SPAARS emphasize the roles, goals and plans of individuals related to the cycles of appraisal. The extent to which an individual experiences emotions depends considerably on the degree of appropriation of roles, on the degree of feasibility of plans and on the conditions of goal achievement. Basic emotions are considered as modules and they can be coupled together. Complex emotions as well as emotional 'disorders' are derived from the same basic emotion set. In SPAARS, the individual is not only aware of his own emotional state, but also aware of his own action potential (Figure 2).

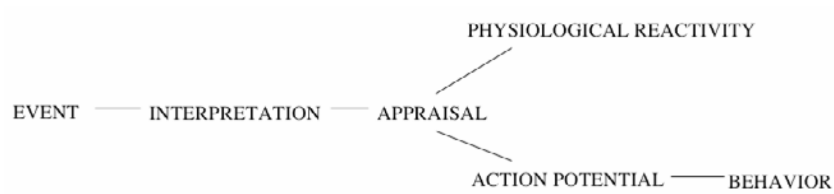


Fig. 2. Appraisal Cycle in SPAARS

3.4 Self-Regulatory Executive Function (*SREF*)

SREF [12] is an integrative model of cognitive-attentional processing which predisposes to emotional distress. The model proposes three levels. In the automated lower level, the elementary processing units are activation-driven, and the resources are domain specific. Processing is unconscious.

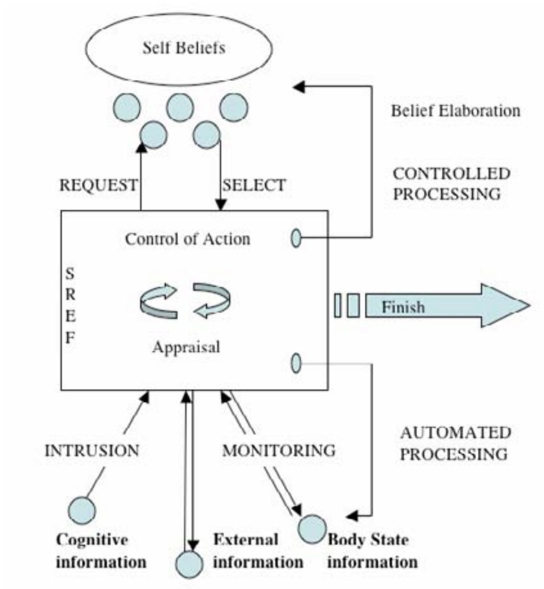


Fig. 3. SREF model

Three types of information can be represented at this level: (1) external stimulus information; (2) cognitive state information; (3) body state information. The second level corresponds to conscious appraisal and regulation of action. The third level consists of acquired knowledge about the self and strategies for self-regulation stored in the long-term memory. SREF performs the appraisal of lower-level information and regulates the action in order to reduce self-discrepancies and perceived threats to the self. The processing is influenced by self-beliefs that impact on both appraisals and strategies.

Top-down, SREF is triggered by conscious strategies and bottom-up by the intrusion of thoughts, sensations and external information into consciousness. Following the intrusion, the significance is appraised and plans of self-regulation are used. Expectancies concerning the ability to succeed depend on situation appraisal and on self-beliefs, and determine the selection of action. Monitoring is centered on self-relevant low-level information.

SREF has been used to explain the impact of negative emotions, i.e. anxiety, on the selective attention according to a plan or strategy to achieve a specific goal. If

the goal is not achieved successfully, then in some cases negative beliefs may intrude either at the automated level of processing, either at the controlled level. There is an impact of negative beliefs if SREF is activated with time-sharing between the plans to achieve the goal and the self-regulative plans. Furthermore, strategies of hypervigilance for threat that increase monitoring or active search for threat may be adopted. Self beliefs have also been mentioned as important in the operational field [1].

4 Integration

The effort of integration is motivated by the structural similarity of the theories presented above. All of them propose several levels, with the distinction between automated processing and controlled processing. All of them deal with goals, and the concepts of Mental Models or Schema are also present in most of them, either defined as such, either defined as cognitive processes. It is interesting to notice that the theories of emotion emphasize clearly the final output as a concrete behavior or action, while the cognitive theories emphasize more on the description of cognitive processes.

Figure 4. illustrates an attempt to put together most of the theories presented above. The bottom-left part under the Self Awareness Situation Awareness dot-line shows the SAGAT model of Situation Awareness, while the top-right part above the Self Awareness Situation Awareness dot-line shows the SREF model of Self Awareness. Labels corresponding to the levels of SRK and SPAARS appear related to their equivalents proposed in the other models. Outputs of the SAGAT model have been connected to inputs of SREF, and the output action of SREF loops back to the SAGAT goal.

In order to keep a neutral unifier for all these theories, we propose as in previous studies multi-agent systems as a flexible and modular structure. Indeed, we find in the Belief, Desire and Intentions (*BDI*) [4] model most of the concepts used in the cognitive and emotion theories.

Beliefs contain the informational state and are only required to provide information on the likely state of the environment. Desires refer to the motivational state, and Intentions refer to the deliberative state.

There are two models in the external viewpoint of BDI: the Agent Model that defines and creates instances of agents and the Interaction Model that defines the interaction and communication between the agents.

There are three models in the internal viewpoint of BDI: the Belief Model that describes the information about the environment and internal state that an agent may hold, and the actions it may perform; the Goal Model that describes the goals that an agent may possibly adopt and the events to which it can respond; the Plan Model that describes the plans that an agent may possibly employ to achieve its goals.

A BDI agent can be completely specified by the events it can perceive, the actions it may perform, the beliefs it may hold, the goals it may adopt, the plans that give rise to its intentions.

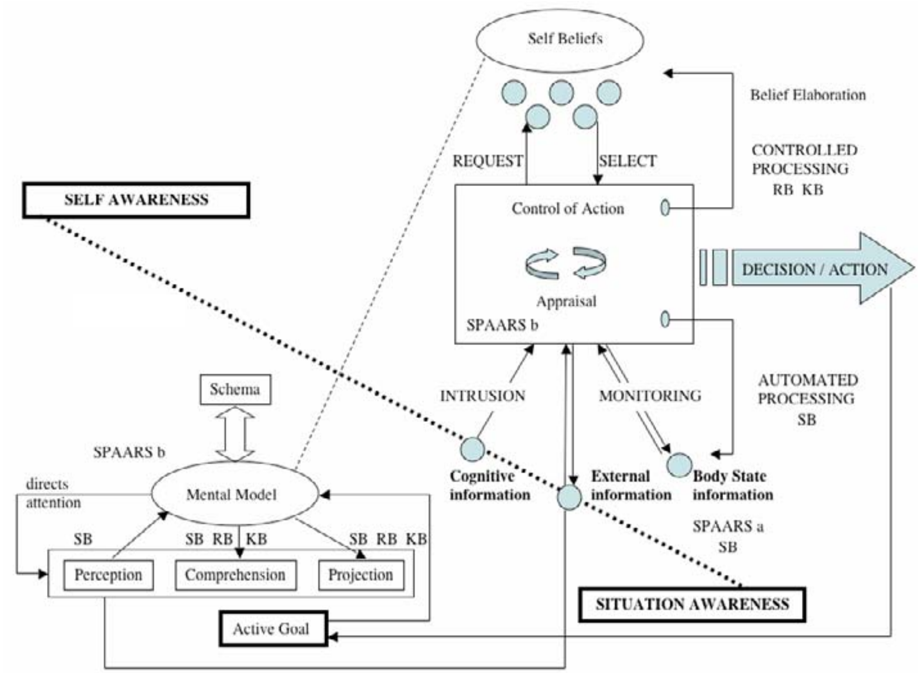


Fig. 4. Situation Awareness and Self Awareness based on SAGAT and SREF diagrams; the other theories are represented by labels: *SB*, *RB*, *KB* stand for Skill Based, Rule Based and Knowledge Based levels; *SPAARS a* stands for the associative level and *SPAARS b* for the Schema level

5 Discussion and Conclusions

The integration effort has to be continued in order to obtain a complete and coherent theory that can provide solutions to assess Workload, Situation Awareness and Self Awareness (emotions). In the end it would be easier to use one theory than a patchwork.

However, in the current state of practice there is still a need to use methods individually, and not necessarily to mix all of them, according to the actual certification documents used by industrials.

The integration is purposeful especially for the design phase as well as for the result analysis step in the evaluation phase, when data and results from different methods have to be correlated. Operational tools for emotion assessment in safety critical industry have to be built and tested.

Self Awareness may profitably contribute to better understand user behaviors, and to finally systematically address emotions in safety critical systems [1].

References

1. Beaty, D.: The naked Pilot. Airlife Publishing Ltd, England (1995)
2. Boy, G.A.: Cognitive Function Analysis. Ablex. Distributed. Greenwood Publishing, Westport (1998)
3. Endsley, M.R.: Designing for situation awareness; An approach to User-Centered Design. Taylor & Francis, London (2003)
4. Kinny, D.: Georgeff, M.: Modeling and Design of Multi-Agent Systems. Technical Note 59. Australian Artificial Intelligence Institute (1996)
5. Lane, R.D., Nadel, L. (eds.): Cognitive Neuroscience of Emotion. Oxford University Press, New York (2000)
6. Power, M., Dalgleish, T.: Cognition and Emotion. From order to disorder. Psychology Press. Taylor & Francis, Hove and New York (1997)
7. Rasmussen, J.: Skills, rules, knowledge: Signals, signs, and symbols and other distinctions in human performance models. IEEE Trans. Syst., Man, Cybern. SMC-13, 257–267 (1983)
8. Richelle, M., Requin, J., Robert, M. (eds.): Traité de Psychologie Expérimentale. vol. 2 (chapitre IV), Presses Universitaires de France, Paris, pp. 3–78 (1994)
9. Schütte, S., Eklund, J.: Product Design for Heart and Soul. An introduction to Kansei Engineering Methodology. Linköpings Universitet, Department for Human Systems Engineering, Uni Tryck Linköping, Sweden (2003)
10. Stephane, L., Boy, G.A.: A Cross-Fertilized Evaluation Method based on visual Patterns and Cognitive Functions. vol. 1 - Engineering Psychology, Health and Computer System Design. HCI International 2005, July 22-27, Las Vegas, NV USA. MIRA Digital Publishing (2005)
11. Taylor, R.M., Finnie, S.E., MacLeod, I.: Enhancing Situational Awareness Through System Cognitive Quality. Defence Research Agency, Farnborough, Hampshire (1996)
12. Wells, A., Matthews, G.: Attention and Emotion. A clinical perspective. Psychology Press. Taylor & Francis, Hove UK (1994)

Sociotechnical Theory and NEC System Design

Guy H. Walker, Neville A. Stanton, Dan Jenkins,
Paul Salmon, Mark Young, and Amerdeep Aujla

Defence Technology Centre for Human Factors Integration (DTC HFI),
Brunel University, BIT Lab, Uxbridge, UB8 3PH

Abstract. The term 'Sociotechnical' is much more than merely a buzzword, it is a set of theories and concepts that seek to jointly optimize the co-evolution of organizations and technology. In this paper we show how the specification of a set of basic sociotechnical principles (responsible autonomy, adaptability and meaningfulness of tasks) could help to create the initial conditions from which shared awareness (through peer to peer interaction) and agility (through Effects Based Operations, semi-autonomous groups and increased tempo) create self-synchronization (synergy, with simple organizations undertaking more complex tasks). Sociotechnical theory has further advantage of a long legacy in theoretical and applied settings, one that future work aims to exploit in order to help realize the vision that NEC promises.

Keywords: Sociotechnical theory, sociotechnical systems, network enabled capability, agility, tempo, shared awareness, effects based operations, organizational design.

1 Introduction

Socio (of people and society) and technical (of machines and technology). Sociotechnical (all one word) or socio-technical (with a hyphen)?[1] Sociotechnical Theory and/or Sociotechnical System? All of these terms appear ubiquitously in HFI literature but their actual meanings often remain unclear.

In its most common usage Sociotechnical refers simply, and quite correctly, to any kind of organisation that is comprised of people and technology. Depending on how technology is defined this, then, could include virtually any or all organisations. There is more to it. Lets, therefore, begin this paper with some orientating sociotechnical facts.

Sociotechnical is much more than merely a buzzword, it is a specific set of theories and concepts that speak towards the interrelatedness of 'social' and 'technical' aspects of an organisation. Sociotechnical 'theory' is founded on two main principles. One is that the interaction of social and technical factors creates the conditions for successful (or unsuccessful) organisational performance. This interaction is comprised partly of

[1] Socio-technical and sociotechnical are identical terms that are used interchangeably in the wider literature. In this paper, the term 'sociotechnical' (without a hyphen between socio and technical) will be adopted.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 619–628, 2007.

linear 'cause and effect' relationships (the relationships that are normally 'designed') and partly from 'non-linear', complex, even unpredictable relationships (the good or bad relationships that are often unexpected). Whether designed or not, both types of interaction occur when technical elements are put to work and socio elements adapt it and themselves to the working context. The corollary of this, and the second of the two main principles, is that optimisation of each aspect alone (socio or technical) tends to increase not only the quantity of unpredictable, 'un-designed' relationships, but those relationships that are actually injurious to the system's performance. Sociotechnical theory, therefore, is about 'joint optimisation' and designing different kinds of organisation, ones in which the relationships between socio and technical elements lead to the emergence of productivity and wellbeing, rather than the all too often case of new technology failing to meet the expectations of designers and users alike.

A sociotechnical 'system', on the other hand, is usually the term given to any instantiation of socio and technical elements engaged in goal directed behaviour. Sociotechnical Systems are a particular expression of sociotechnical theory (although they are not necessarily one and the same thing). The term sociotechnical 'system' recognises that organisations have boundaries and that transactions occur within the system (and its sub-systems) and between the wider context and dynamics of the environment. It is an extension of sociotechnical theory, one that provides a richer descriptive and conceptual language for describing, analysing and designing organisations based on General Systems Theory.

Sociotechnical theory (which subsumes sociotechnical systems) provides a rich conceptual language, as well as specific organization design principles, both of which are isomorphic with the long term vision of Network Enabled Capability (NEC) in military settings. The purpose of this paper is to demonstrate the numerous ways in which sociotechnical theory and NEC concepts overlap.

2 Origins

Sociotechnical theory and NEC are a response to a particular form of organizational design, the stereotypical case of a rigidly hierarchical bureaucracy of which notions of Classic C2 are an example. NEC is a response to the increasing complexity and dynamism of the military landscape, to which Classic C2 is becoming an increasingly inefficient organizational paradigm. Classic C2 is characterized by unitary decision rights (in which optimum means to ends are specified at the top of, or at higher levels of a vertical hierarchy); tightly constrained patterns of interaction (due to rules, standard operating procedures and other means by which the bureaucracy embodies its past experience and specifies optimum means to ends) and tight control (in which performance can be quantified and controlled through intermediate echelons of management; NATO, 2006). In other words, Classic C2 is a highly 'designed' form of organization.

Classic C2 rests on a set of deeper, more fundamental presuppositions. It can be seen as "…the most efficient structure for handling large numbers of tasks…no other structure could handle the massive quantity of work as efficiently" (Ritzer, 1993, p.20). It is also promises predictability: "Outsiders who receive the services [that

Classic C2 dispenses] know with a high degree of confidence what they will receive and when they will receive it" (Ritzer, 1993, p. 21). The performance of Classic C2 organisations is also quantifiable, it is able "to express [its] achievements in terms of activity rather than effect. That is, deploying statistics of missions flown and bombs dropped" (Storr, 2005, p. 34) in which 'more' is considered excellence and less is substandard. Finally, Classic C2 emphasises control: "…[it] may be seen as one huge nonhuman technology. Its nearly automatic functioning may be seen as an effort to replace human judgement with the dictates of rules, regulations and structures" (Ritzer, 1993, p. 21). Having said all of that, we fully acknowledge the stereotypical nature of this characterization, yet many of these features will be recogniseable in essence if not extent. The fundamental irony still stands:

"Instead of remaining efficient, bureaucracies can degenerate into inefficiency as a result of 'red tape' and the other pathologies we usually associate with them. Bureaucracies often become unpredictable as employees grow unclear about what they are supposed to do and clients do not get the services they expect. The emphasis on quantification often leads to large amounts of poor-quality work…All in all, what were designed to be highly Rational operations often end up growing quite irrational" (Ritzer, 1993, p.22).

This is why the paradigm shift represented by Sociotechnical theory and, latterly in the case of the military, NEC, is currently underway. In the specific case of NEC the shift is made possible, indeed, it is motivated by the capabilities of emerging technology, in particular the distributed, ubiquitous nature of networking technologies such as the internet. Thus in one sense technology and organizations can be seen to be 'co-evolving'; new types of organization place new requirements on technology, in return, new technology creates new possibilities for organisations (Alberts, Gartska and Stein, 2005). The transit of this co-evolution is anything but smooth. Quite often the state of 'technology push' is created, that is, technology is ahead of the type of organizations that could benefit from it most. Arguably, this is where the incipient state of NEC's development is at the current time; communications and computing technology are effectively being bolted on to an existing 'type' of organisation in an attempt to create a kind of 'high speed hierarchy'.

The alternative vision is a jointly optimized solution in which the co-evolution of organizations and enabling technology is fully exploited rather than just one or the other. This is where the concept (if not the reality) of NEC is properly located. This relates to the now well worn maxim that NEC is much more than just a technology, it is also much more than Classic C2.

3 Network Enabled Capability

It is probably fair to say that NEC is at once a frequently over-defined term and, yet, still relatively ill-defined. The central tenets of the prevailing vision of NEC, in terms of what this type of organization aims to deliver, seem to be distillable down to the following:

Shared awareness *"includes the sharing not only of information on the immediate environment and intentions of our own enemy and neutral forces, but also the development of shared combat intent and understanding"* (Ferbrache, 2005, p. 104).

Specific instantiations of this feature include (but are not confined to) ideas about joint operational pictures and issues concerning distributed collaborative planning.

Agility, the *“ability to reconfigure forces and structures rapidly, building on this shared awareness, exploiting effective mission planning methods, and enabled by an information environment that allows rapid reconfiguration of the underlying network and knowledge bases”* (Ferbrache, 2005, p. 104).. This encompasses specific issues such as mission-organised force elements, effects-based operations and enhanced tempo.

Synchronization, the *“ability to plan for and execute a campaign in which we can ensure all elements of the force work together to maximum military effect by synchronizing the execution of their missions to mass forces or generate coordinated effects on target”* (Ferbrache, 2005, p. 104).

The techno-organisational vision of NEC, therefore, runs along the following lines:

“…self-synchronizing forces that can work together to adapt to a changing environment, and to develop a shared view of how best to employ force and effect to defeat the enemy. This vision removes traditional command hierarchies and empowers individual units to interpret the broad command intent and evolve a flexible execution strategy with their peers” (Ferbrache, 2005, p. 104).

Although there is little evidence of overt cross-referencing, this vision is shared almost exactly by sociotechnical theory (with perhaps less emphasis on ‘technology’ as we would interpret it today). However, sociotechnical theory, unlike NEC, has the advantage of a 50 year pedigree of organizations in which such a vision has been realized in practice with considerable success (e.g. Pasmore et al., 1982; Beekun, 1989; Cummings, Molloy & Glen, 1977). In the remainder of the paper we will demonstrate how sociotechnical theory proposes to achieve synchronization and, by implication, the promised vision of NEC through the joint optimization of technology and organizational design.

4 Key Sociotechnical Characteristics

Some of the central principles of sociotechnical theory were elaborated in a seminal paper by Trist and Bamforth in 1951. This is an interesting case study which, like most of the work in sociotechnical theory, is focused on a form of ‘production system’ expressive of the era and the contemporary technological systems it contained. This was an era long before technology made concepts like NEC a reality (or even a possibility); yet the parallels are immediately apparent and no less relevant.

The study was based on the paradoxical observation that despite improved technology, productivity was falling, and that despite better pay and amenities, absenteeism was increasing. This particular rational organisation had become irrational. The cause of the problem was hypothesized to be the adoption of a new form of production technology which had created the need for a bureaucratic from of organization (rather like classic C2). In this specific example, technology brought with it a retrograde step in organizational design terms. The analysis that followed

introduced the terms ‘socio’ and ‘technical’ and elaborated on many of the core principles that sociotechnical theory subsequently became.

4.1 Responsible Autonomy

Sociotechnical theory was pioneering for its shift in emphasis, a shift towards considering teams or groups as the primary unit of analysis and not the individual. Sociotechnical theory pays particular attention to internal supervision and leadership at the level of the ‘group’ and refers to it as ‘responsible autonomy’ (Trist & Bamforth, 1951, p.6). NEC, on the other hand, refers to it as peer-to-peer interaction. The overriding point seems to be that having the simple ability of individual team members being able to perform their function is not the only predictor of combat effectiveness. There are a range of issues in team cohesion research, for example, that are answered by having the regulation and leadership internal to a group or team (e.g. Siebold, 2000; Oliver, et al., 2000; Mael & Alderks, 2002). These, and other factors, play an integral and parallel role in ensuring successful teamwork which sociotechnical theory exploits.

The idea of semi-autonomous groups conveys a number of further advantages. Not least among these, especially in hazardous military environments, is the often felt need on the part of people in the organisation for a role in a small primary group. It is argued that such a need arises in cases where the means for effective communication are often somewhat limited. As Carvalho (2006) states, this is because “…operators use verbal exchanges to produce continuous, redundant and recursive interactions to successfully construct and maintain individual and mutual awareness…”. In Classic C2 terms the immediacy and proximity of trusted team members makes it possible for this to occur. In the distributed world of NEC, the co-evolution of technology and organizations brings with it an expanding array of new possibilities for novel interaction. Responsible autonomy could become more distributed along with the team(s) themselves.

The key to responsible autonomy seems to be to design an organization possessing the characteristics of small groups whilst preventing the ‘silo-thinking’ and ‘stovepipe’ neologisms of contemporary management theory. In order to preserve “…intact the loyalties on which the small group [depend]…the system as a whole [needs to contain] its bad in a way that [does] not destroy its good” (Trist & Bamforth, 1951, p. 9). In practice (e.g. Rice, 1958) this requires groups to be responsible for their own internal regulation and supervision, with the primary task of relating the group to the wider system falling explicitly to a group leader. This principle, therefore, describes a strategy for removing more traditional command hierarchies.

4.2 Adaptability

Carvajal (1983) states that “the rate at which uncertainty overwhelms an organisation is related more to its internal structure than to the amount of environmental uncertainty”. Sitter, Hertog and Dankbaar (1997) offer two solutions for

organisations confronted, like the military, with an environment of increased (and increasing) complexity:

"The first option is to restore the fit with the external complexity by an increasing internal complexity. […] This usually means the creation of more staff functions or the enlargement of staff-functions and/or the investment in vertical information systems" (p. 498). Vertical information systems are often confused for NEC but an important distinction needs to be made, which Sitter et al propose as their second option:

"…the organisation tries to deal with the external complexity by 'reducing' the internal control and coordination needs. […] This option might be called the strategy of 'simple organisations and complex jobs'". This all contributes to a number of unique advantages.

Firstly is the issue of 'human redundancy' (e.g. Clarke, 2005) in which "groups of this kind were free to set their own targets, so that aspiration levels with respect to production could be adjusted to the age and stamina of the individuals concerned" (Trist & Bamforth, 1951, p. 7). Human redundancy speaks towards the flexibility, ubiquity and pervasiveness of resources within NEC.

The second issue is that of complexity. Complexity lies at the heart of Classic C2's shortcomings (it is an organizational paradigm that struggles to cope with it) and at the heart of NEC's manifold implementation issues (complexity makes it difficult to implement the required technological infrastructure in a similar way to that already found in civilian domains). Trist and Bamforth (1951) could have been writing about the military context with this passage: *"A very large variety of unfavourable and changing environmental conditions is encountered […], many of which are impossible to predict. Others, though predictable, are impossible to alter."* (Trist & Bamforth, 1951, p.20). Classic C2 is clearly motivated by the appealing 'industrial age', rational principles of 'factory production', a particular approach to dealing with complexity: "In the factory a comparatively high degree of control can be exercised over the complex and moving 'figure' of a production sequence, since it is possible to maintain the 'ground' in a comparatively passive and constant state" (Trist & Bamforth, 1951, p. 20). On the other hand, military activities are constantly faced with the possibility of "untoward activity in the 'ground'" of the 'figure-ground' relationship" (Trist & Bamforth, 1951, p. 20). The central problem, one that appears to be at the nub of Classic C2's shortcomings, is that "The instability of the 'ground' limits the applicability […] of methods derived from the factory" (Trist & Bamforth, 1951, p. 20).

In Classic C2 problems with the moving 'figure' and moving 'ground' often become magnified through a much larger social space, one in which there is a far greater extent of hierarchical task interdependence (Trist & Bamforth, 1951, p.21). For this reason, the semi-autonomous group, and its ability to make a much more fine grained response to the 'ground' situation, can be regarded as 'agile'. Added to which, local problems that do arise need not propagate throughout the entire system (to affect the workload and quality of work of many others) because a complex organization doing simple tasks has been replaced by a simpler organization doing more complex tasks. The agility and internal regulation of the group allows problems to be solved locally without propagation through a larger social space, thus increasing tempo.

Another concept in sociotechnical theory is the 'whole task'. A whole task "has the advantage of placing responsibility for the […] task squarely on the shoulders of a single, small, face-to-face group which experiences the entire cycle of operations within the compass of its membership." (Trist & Bamforth, 1951, p. 6). The Sociotechnical embodiment of this principle is the notion of minimal critical specification. This principle states that, "While it may be necessary to be quite precise about what has to be done, it is rarely necessary to be precise about how it is done" (Cherns, 1976, p. 786). This is no more illustrated by the antithetical example of 'working to rule' and the virtual collapse of any system that is subject to the intentional withdrawal of human adaptation to situations and contexts.

The key factor in minimally critically specifying tasks is the responsible autonomy of the group to decide, based on local conditions, how best to undertake the task in a flexible adaptive manner. This principle is isomorphic with Effects Based Operations (EBO). EBO asks the question of what goal is it that we want to achieve, what objective is it that we need to reach rather than what tasks have to be undertaken, when and how. The EBO concept enables the commander to "…manipulate and decompose high level effects. They must then assign lesser effects as objectives for subordinates to achieve. The intention is that subordinates' actions will cumulatively achieve the overall effects desired" (Storr, 2005, p. 33). In other words, the focus shifts from being a scriptwriter for tasks to instead being a designer of behaviours. In some cases this can make the task of the commander significantly less arduous (e.g. Rice, 1958).

4.3 Meaningfulness of Tasks

Effects Based Operations and the notion of a 'whole task', combined with adaptability and responsible autonomy, have additional advantages for those at work in the organization. This is because "for each participant the task has total significance and dynamic closure" (Trist & Bamforth, 1951, p. 6) as well as the requirement to deploy a multiplicity of skills and to have the responsible autonomy in order to select when and how to do so. This is clearly hinting at a relaxation of the myriad control mechanisms found in organizations like Classic C2.

Multi force and multi-national operations also bring with them an issue of size, in which "the scale of a task transcends the limits of simple spatio-temporal structure. By this is meant conditions under which those concerned can complete a job in one place at one time, i.e., the situation of the face-to-face, or singular group" (Trist & Bamforth, 1951, p. 14). In other words, in Classic C2 the 'wholeness' of a task is often diminished by multiple group integration and spatiotemporal disintegration (Trist & Bamforth, 1951, p. 14). The group based form of organization design proposed by sociotechnical theory combined with the technological possibilities of NEC provides a response to this often forgotten issue, one that contributes significantly to joint optimisation.

5 The Alternative Conception of a 'Network' in NEC

The phrase Sociotechnical 'System' is even more ubiquitous than merely 'Sociotechnical'. Sociotechnical Systems is a particular instantiation of sociotechnical theory which uses the language, metaphors and concepts of general system theory

(e.g. Bertalanffy, 1950). It helps us to explain exactly why responsible autonomy, adaptability and meaningfulness of tasks render the emergence of highly desirable properties like shared awareness, agility and synchronization.

An important distinction needs to be made between closed and open systems. Closed systems, also referred to as 'Rational systems' (Scott, 1992), have the following characteristics: they are typically concerned with the attainment of a relatively specific goal, there are often well specified criteria for deciding on optimum means to ends, there is "a relatively high degree of formalization" and there is "conscious and deliberate" cooperation among participants (i.e. the organisation is 'designed'). In closed systems the boundary is drawn around the technical, functional and structural elements of a work domain and the links between them are, and indeed can be, tightly defined. In systems parlance we are to some extent no longer referring to a 'network' at all (a more conceptual way of linking entities together in which the linkage between entities is loosely specified) instead we are talking of an actual 'object' (in which information flows are known and definable). Classic C2 could perhaps be regarded as more of an object. NEC, on the other hand, is a particular type of 'open system':

"The alternative conception of 'open systems' carries the implication that such systems may spontaneously re-organise towards states of greater heterogeneity and complexity, and that they achieve a 'steady state' at a level where they can still do work. Enterprises would appear to possess such 'open systems' characteristics. They grow by processes of internal elaboration. They manage to achieve a steady state while doing work. They achieve a quasi-stationary equilibrium in which the enterprise as a whole remains constant, with a continuous 'throughput', despite a considerable range of external changes." (Bertalanffy, 1950, p. 45).

Responsible autonomy, adaptability and meaningfulness of tasks create the conditions for this appealing type of system to 'emerge'. They create the conditions from which "a final state may be reached from different initial conditions and in different ways" (Bertalanffy, 1950, p. 25), in other words, a flexible execution strategy, which is precisely what seems to be required when an organization is confronted, like the military, with a complex adaptive environment.

An alternative conception of NEC is that it is specified as much by the linkages between parts as the parts themselves, in other words "the set of relations determines the very character of the system…[and]…the structure of the system determines its function" (Ropohl, 1999, p. 4). If NEC is defined by sociotechnical principles, then according to Metcalf's Law, as the number of NEC's 'parts' increase incrementally its effectiveness as a whole should increase exponentially. The network becomes synchronous, more than the sum of its parts, it becomes synergistic. From a Sociotechnical perspective NEC seems to be as much about creating the conditions for networks to become more than the sum of their parts than it is about highly complex formal architectures or similarly complex technology.

6 Conclusions

Sociotechnical is much more than merely a buzzword, it is a set of theories and concepts that seeks to jointly optimize the co-evolution of organizations and

technology. This paper has briefly shown how the specification of a set of basic sociotechnical principles (responsible autonomy, adaptability and meaningfulness of tasks) could help to create the initial conditions from which shared awareness (through peer to peer interaction) and agility (through EBO, semi-autonomous groups and increased tempo) create self-synchronization (synergy and simple organizations doing complex tasks).

Sociotechnical theory is uniquely placed. This paper would be entirely hypothetical, like a considerable portion of current NEC research, were it not for a 50 year legacy of organizational design according to these principles combined with an overwhelming track record of real-world success. Sociotechnical theory, unlike some of the more analytical approaches applied to NEC, is also a very human centered theory. It is concerned as much with the optimization of ‘productivity’ as it is with the experiences of people working in organizations. This overriding principle of joint optimization is a key factor in its success and especially concordant with the aims and objectives of the DTC HFI.

To conclude, sociotechnical theory is a unique selling point of the DTC HFI’s approach to NEC system design. Current and future work is directed towards turning theory into practice, for which, we argue, sociotechnical theory is uniquely positioned.

References

1. Alberts, D.S., Garstka, J.J., Stein, F.P.: Network centric warfare: Developing and leveraging information superiority. CCRP (1999)
2. Beekun, R.I.: Assessing the effectiveness of sociotechnical interventions: Antidote or fad? Human Relations 42(10), 877–897 (1989)
3. v Bertalanffy, L.: The theory of open systems in physics and biology. Science 111, 23–29 (1950)
4. Carvajal, R.: Systemic netfields: the systems’ paradigm crises. Part I. Human Relations 36(3), 227–246 (1983)
5. d Carvalho, P.V.R.: Ergonomic field studies in a nuclear power plant control room. Progress in Nuclear Energy 48, 51–69 (2006)
6. Cherns, A.: The principles of sociotechnical design. Human Relations 29(8), 783–792 (1976)
7. Clark, D.M.: Human redundancy in complex, hazardous systems: A theoretical framework. Safety Science 43, 655–677 (2005)
8. Cummings, T.G., Molloy, E.S., Glen, R.: A methodological critique of fifty-eight selected work experiments. Human Relations 30(8), 675–708 (1977)
9. Ferbrache, D.: Network enabled capability: concepts and delivery. Journal of Defence Science 8(3), 104–107 (2003)
10. Mael, F.A., Alderks, C.E.: Leadership team cohesion and subordinate work unit morale and performance. Military Psychology 5(3), 141–158 (2002)
11. NATO (2006). SAS-050 conceptual model version 1.0. Available at: http://www.dodccrp.org/SAS/SAS-050%20Final%20Report.pdf
12. Oliver, L.W., Harman, J., Hoover, E., Hayes, S.M., Pandhi, N.A.: A quantitative integration of the military cohesion literature. Military Psychology 11(1), 57–83 (2000)

13. Pasmore, W., Francis, C., Haldeman, J., Shani, A.: Sociotechnical systems: a north American reflection on empirical studies of the seventies. Human Relations 35(12), 1179–1204 (1982)
14. Rice, A.: Productivity and social organisation: The Ahmedabad experiment. London: Tavistock (1958)
15. Ritzer, G.: The McDonaldization of society. Pine Forge Press, London (1993)
16. Ropohl, G.: Philosophy of socio-technical systems. Society for Philosophy and Technology, 4(3). On-Line article accessed on 11th August 2006 (1999)
17. Scott, R.: Organizations; Rational, natural, and open systems. Prentice Hall, New Jersey (1992)
18. Siebold, G.L.: The evolution of the measurement of cohesion. Military Psychology 11(1), 5–26 (2000)
19. Sitter, L.U., Hertog, J.F., Dankbaar, B.: From complex organizations with simple jobs to simple organizations with complex jobs. Human Relations 50(5), 497–536 (1997)
20. Storr, J.: A critique of effects-based thinking. RUSI Journal (2005)
21. Trist, E., Bamforth, K.: Some social and psychological consequences of the longwall method of coal getting. Human Relations 4, 3–38 (1951)

Part IV

Safety Critical Applications and Systems

The Impact of Automation and FMS in Flight Safety: Results of a Survey and an Experimental Study

Diyar Akay[1], Ergün Eraslan[2], and Cengiz Yoldaş[3]

[1] Gazi University, Department of Industrial Engineering, Ankara, Turkey
diyar@gazi.edu.tr
[2] Başkent University, Department of Industrial Engineering, Ankara, Turkey
eraslan@baskent.edu.tr
[3] Commercial Pilot, Erler Mah, Ozbay Apt., Etimesgut, Ankara, Turkey
cengoyol@yahoo.com

Abstract. The main aim in the adoption of automation technologies of airplanes is to decrease the errors due to human factors whereas some applications and empirical researches showed that automation causes new type of errors in the cockpit. When the aviation accidents due to automation have been surveyed, it is seen that the automation errors related to FMS (Flight Management System) are very high. The FMS supports pilots in different tasks like flight planning, navigation, performance management and monitoring of flight progress. In this study, the effect of automation and FMS in flight safety is researched. A survey was conducted to a group of pilots of a turboprop airplane with high level automation. Following the survey, an experimental study was performed to analyze effects of last minute changes on FMS which was highlighted as the most prominent problem in the survey. Finally, based on the results of experiment, recommendations were given so as to improve design and usability of FMS.

Keywords: Automation Errors, Cockpit Automation, Flight Management System.

1 Introduction

Automation can be defined as the accomplishment of a job either in a mechanical or an electrical way. In most of the cases, automation carries out the things which can be done by humans however; in some cases it accomplishes many things that pass the humans' cognitive and physical limits [1]. In recent years, a sharp rise is observed that automation has started to be used in aviation as well as in many different areas. Especially in flight decks, there has been a swift movement to the usage of digital systems instead of mechanical systems.

In addition to the aims like, increasing cockpit automation level, decreasing crew workload, making the system work in a better and more reliable way, the major aim of automation is to minimize the errors due to human factors as well as the aviation accidents as a result of these errors. According to the researches, approximately 73% of the accidents occur due to human factors, and CFIT (Controlled flight into terrain)

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 631–638, 2007.

accidents are the most common ones in today's world [2]. The main aim in the adoption of automation technologies of airplanes is to decrease the errors due to human factors whereas some applications and empirical research showed that automation causes new type of errors in the cockpit [3]. Besides, Billings suggest that technological advances, incorporated in the modern cockpit have had the effect of making the flight crew more peripheral to the actual operation of the aircraft [4].

Cockpit automation has had many positive effects on flight safety. For instance, it has increased technical safety and effectiveness and decreased costs. Besides, with the aid of automation, cockpits were simplified, situational awareness of the pilots increased, crew workload during some period of the flight decreased and airplanes were made to have a smoother and more accurate flight controls [2]. As a result, when it is seen from this point of view, automation has positive effects on flight safety. But, on the other hand, it has some negative effects. In a long range airplane with a high automation level, after long cruise hours, it becomes possible for the pilots to lose situational awareness due to hypovigilance. Hard to detect errors, automation complacency or mistrust in automation, confusion, blinkered concentration, hard to learn complex systems, mode errors, being out of the loop and loosing manual flying skills, more heads down time, difficulties with the crew co-ordination and communication difficulty in changing plans are the other disadvantages of automation [2].

Computers and decision support systems were put into cockpit to decrease human errors and help the pilot but it is thought that, users perceive these systems wiser and superior. Depending on those psychological factors, pilots, in need of making a decision during the flight, can trust on automation without considering all the data. This is called automation bias. As a result of considering automation an authority, "omission errors" and "commission errors" occur [5]. It is also possible to categorize the errors by their source. An automation error can occur due to a failure of the system or a function. Improper and incorrect programming of the system done by the user or a false input to the system can cause automation error. Other causes of errors may be inadequate training and unsuitable operating procedures. Finally, the errors made at the stages of designing and programming of software systems can cause automation errors [6].

When the aviation accidents, which automation takes part in, are surveyed, it is seen that the automation errors related to FMS are very high [7]. Today FMS is used widely in large airliners as well as in light general aviation airplanes. The FMS support pilots in different tasks like the flight planning, navigation, performance management and monitoring of flight progress [8]. In this research, FMS, in other words, the part of automation in cockpit with which the pilot is interacting mostly is studied widely. A survey is given to a group of pilots of a turboprop airplane with high level of automation and the results of the survey are used in a flight simulator based experiment to analyze effects of the ATC originated last minute changes made on FMS. Finally, based on the results of experiment, recommendations were given so as to improve design and usability of FMS.

2 Survey

A group of pilots flying with high automation level airplanes (B-200) were selected for the survey. The most important feature of these pilots was their experience of

flying with conventional cockpit planes before B-200. Transferring into B-200 means; having met an autopilot and digital displays for the first time. EFIS (Electronic Flight Instrumentation System), FMS, and FCS (Flight Control System) in the B-200 has adequate level to do a research on pilot-automation interaction.

Survey was sent to 50 pilots and 34 of them were responded. Return rate of the survey was 68%. The survey was formed based on BASI in Australia and FAA in USA [9]. Survey has 3 parts. Part 1 consists of questions about pilots' status, age, total flight hours etc. Part 2 has 44 questions. These ones are grouped according to Air Traffic Control (ATC), design, mode errors, automation complacency, situational awareness, workload, CRM (Crew Resource Management) and training. Likert Scale was used for this part. In Part 3, 5 open-ended questions were asked to add responders' point of view. Answers in the responded surveys were transferred into a numerical data and analyzed using SPSS.

Part 1: 82% of the pilots who answered the questions were 40 year-old and above. Total flight hours change between 500 and 4500. 100% of them met EFIS and FMS first time on this airplane and 71% of them had their first experience on autopilot with a B-200. It is an important fact that pilots transfer from conventional cockpit to digital cockpit by B-200. 77% of the pilots stated that they faced automation errors at least once a year.

Part 2: In the questions related to ATC and automation, 62% of the pilots confirmed that due to the last minute changes that ATC gave, it took a long time to reprogram systems and especially FMS. When the pressure of obeying the instruction which ATC gave, time limitation and complexity of automation systems coupled, different type errors occurred. Another question in the survey elicited that especially under 10.000 feet in terminal areas, cockpit crew workload increases due to FMS programming and changes in the mode. In the questions about automation design and pilot interaction, 59% of the pilots indicated that they met automation surprises during the flight and 29% of them mentioned that sometimes autopilot can not catch the preselected altitude. 60% of them believed that automation decreases manual skills.

When the answers about automation system modes are examined, the result was; FMS, EFIS and FCS modes causes the pilots increased head-down time therefore outside scan could be obstructed. Also 70% of the pilots experienced mode confusion. 33% of them indicated that they still had no information about the function of some modes. 77% of the responders specified that some mode buttons and mode panels were not user friendly. It causes pilots to bend down to do their work since FMS, EFIS and FCS mode panels are on the lower pedestal. Therefore, 73% of the pilots had concern to lose their orientation while FMS programming and selecting mode.

54% of the pilots found some parts of the automation system unreliable. This result may have been related to their common experience in having encountered automation errors before. 62% of the pilot believed that everything needed for a safe flight is found in FMS database. This shows the high expectation from FMS.

Owing to the past experience with conventional cockpits, it was monitored that an extra effort was given to sustain the situational awareness in automated cockpit.

94% of the pilots pointed out that their workload showed a decline after automation. This claim is true for cruise phase however; during approach and departure stages the workload increases. Another important result from the answers given to CRM questions is; it is emphasized that FMS, EFIS, and FCS related inputs are not easily monitored by the other crew member. 65% of the pilots said that they are willing to learn the automation system deeply. This result confirms the need of reconsidering training curriculum in order to give pilots a full sense of control over automation.

Part 3: When the open-ended questions are evaluated, it was found out that, pilots often encountered automation errors, had problems in user-interface design, and learning FMS especially at initial training stage.

Result: When the survey result is evaluated as a whole, it is inference that pilot automation problems in B-200 airplane are mostly FMS and mode panel centered. FMS has become a device which can not be given up in modern flight decks and today, pilot interaction with FMS has attained great importance. Particularly, this interaction is more critical in the course of last minute changes.

3 Experimental Study

The aim of scenario based experiment is to confirm the result of the survey by creating the unexpected injections (last minute changes) during flight. Outcome of the survey had stressed that the last minute changes given by ATC is causing sudden rise in cockpit workload level and increasing the risk of making faults. Also, due to the need of reprogramming FMS which is located on the lower pedestal, pilot could not pay attention to primary flight instruments and outside scan. In this experiment, a real flight duty has been simulated to observe pilots' behaviors and reaction times. 10 pilots among the responders were selected considering their flight experiences. Initially, a briefing about the scenario was given to the group. Nothing was mentioned about injections because it could affect the result of the ATC changes during normal flight. During the 30 mins experiment in the flight simulator for each pilot, a flight instructor acted as ATC and pilot behaviors were observed. Another person recorded the behaviors and reaction times. In fig. 1 below, the flight route and the injections were given.

According to the scenario, descending airplane for Esenboğa enters Ankara TMA via TELVO waypoint and normally comes to ANK NBD after HAY NBD. Then, it executes ILS 03R approach. During this flight, 4 injections will be given unexpectedly. During these injection periods, pilot's control over FMS and adaptation times to the changes will be observed.

- In the first injection, instruction is given to the pilot to hold over HAY due to separation requirements.
- In the second injection, instruction is given to the pilot to cancel holding and proceed as planned.

- In the third injection, instruction is given to the pilot to program automation and FMS for ILS approach Esenboğa Runway 03R.
- In the fourth injection instruction is given to the pilot to stop descending at 8000 feet and proceed BUK VOR for ILS 21. Then, reprogramme automation and FMS for the new situation.

Results obtained from experiment: Results related to behaviors in the injections are categorized as SATISFACTORY (S), PARTIALLY SATISFACTORY (PS) AND UNSATISFACTORY (U). Results of timings were recorded according to the measurement of time between taking instruction and accomplishment. The experimental results are shown in Table 1.

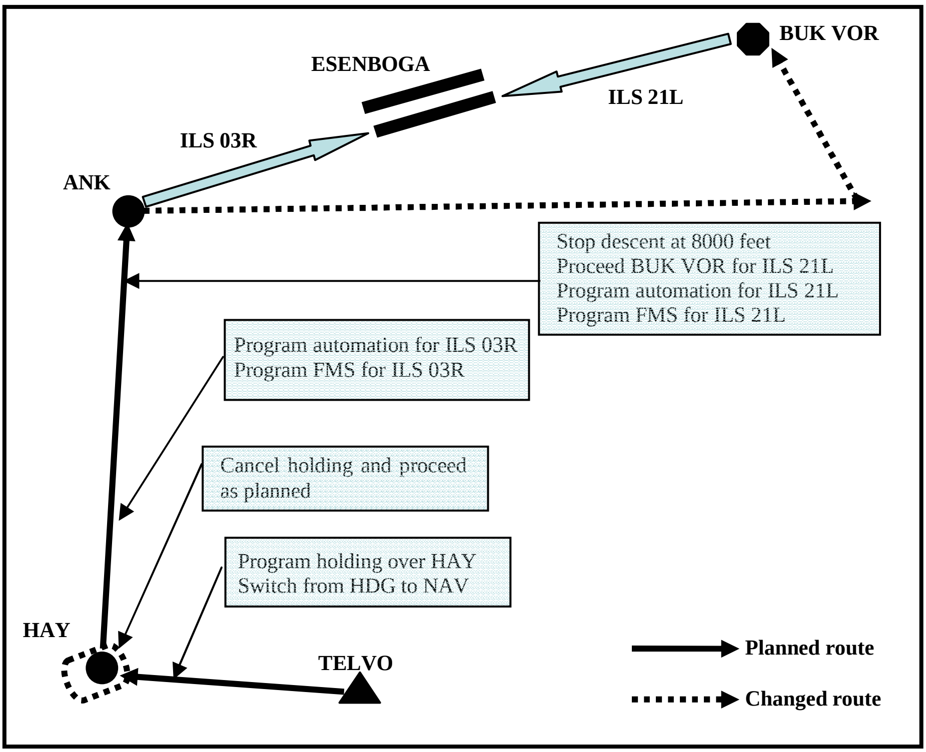


Fig. 1. Planned route and injections for experimental study

In the first injection, there is somewhat lower work load in the cockpit. It was expected from pilots to plan a holding over HAY waypoint. Although there was time constraint due to the short distance to HAY, all pilots completed this duty successfully. Average 33 second was spent for FMS programming to arrange a holding over HAY.

Table 1. Results of the experimental study

INJECTION	TASKS	PILOT-1	PILOT-2	PILOT-3	PILOT-4	PILOT-5	PILOT-6	PILOT-7	PILOT-8	PILOT-9	PILOT-10
1	Hold over HAY	S	S	S	S	S	S	S	S	S	S
	FMS programming time (sec)	40	30	16	55	34	25	24	47	34	25
	Switching from HDG to NAV	S	S	S	S	U	S	S	S	U	S
2	Cancel holding and proceed as planned	PS	PS	S	U	PS	S	PS	PS	S	PS
	Duration (sec)	15	10	11	_	10	10	17	23	13	11
	Method used to cancel holding?	HDG	DIRECT TO	PROCEED	U	DIRECT TO	PROCEED	DIRECT TO	DIRECT TO	PROCEED	DIRECT TO
3	Program automation for ILS 03R	S	S	S	U	S	PS	PS	S	PS	S
	FMS programming	PS	S	S	U	S	S	S	S	PS	S
	FMS programming time (sec)	2’	1’ 30	33	_	55	42	1’ 15	1’ 25	2’ 18	1’ 04
	Omitted items	MKR BCN, DH	MKR BCN, DH	_	U	_	LOC not activ.	–	MKR BCN, DH	MKR BCN, DH	MKR BCN
4	Briefing for ILS 21L	U	U	U	U	U	S	U	U	U	U
	Time for reprogramming automation (sec)	1’	25	1’	_	40	35	1’ 35	35	1’ 24	45
	Stopping descent at 8000 feet	S	S	S	S	S	PS	S	S	S	S
	Time for reprogramming FMS for ILS 21L (sec)	U	3’20	42	U	2 ‘ 10	2’ 20	2’ 50	U	U	1’ 48

Satisfactory (S), Partially Satisfactory (PS), Unsatisfactory (U).

Workload of pilots is in the minimum level as the airplane is in the cruising flight even though instruction was given in a point very close to HAY. As a result, pilots completed the instruction correctly in a short time period.

Another task to be completed by pilots in first injection is to pass from HDG mode to NAV mode after holding planning. Two pilots did not manage this task. Success rate is %80 (Table 1). In case switching from HDG to NAV is not carried out, autopilot will not start holding automatically and plane will continue to fly in a constant heading.

In the second injection, it is demanded to cancel HAY holding by “PROCEED” function and then continue to planned route. One pilot did not accomplish the task. 3 pilot used accurate function (“proceed”) while six pilot followed up trial-error procedures (“direct to or “HDG”) to cancel waiting. This finding show that cognitive model of how to use FMS does not formed in detail.

In the third injection, six of the pilots set up the program automation properly while three of them completed the settings with some shortcomings, and one pilot was characterized as unsatisfactory. In the case of FMS programming, seven pilots completed the task faultlessly, two of them fulfilled the programming with the help of the instructor and, one pilot was unsuccessful. Most part of the automation errors intensify in course transfer, MKR BCN and DH settings.

In the fourth injection, flight process going on in its normal way so far came into a critical phase in consequence of a last minute changes given by ATC. In this injection, 9 pilots made a procedure error by not initiating approach brifing for the new situation. It took about 53 second to program automation which was accomplished by nine pilots. Four pilots were unable to reprogram FMS automation for the new approach plan; on the other hand, it took very long time for the six succesful pilots to reprogram FMS. Approximately 2’ 12 second wasted for 21L ILS programming in FMS in the 4th injection. It is noticed that there is about %70 time increase in the 4th injection compared to FMS programming time in the 3rd injection. A sudden and critical last minute change lead to time pressure on pilots although functions of FMS programming have similar properties in the 3rd and 4th injections.

All these data shed light on 3 results. Firstly, some pilots could not have a mind map of function of autopilot and automation in FMS. Sometimes pilots use trial-error methods or look up pages instead of using a short cut function. This causes an unnecessary workload in the cockpit. The second result is the change in period and accuracy of procedure at normal and critical stages of flight. At normal flight stage, FMS procedure can be finished using a correct method in a short time. When the pilot is enforced for risky procedures at certain stages of the flight (like emergency or ATC instruction change), they spend more time on automation system management. Also the timing pressure on the crew causes errors. Third result is about procedures. When the timing pressure increases, pilots go out of the standard procedures like omitting approach brifing.

4 Conclusions

In this study, an experimental research is carried out to understand the effect of FMS automation system on flight safety. Scenario related results of the experiment showed that, especially last minute changes in airplanes with automation, cause a sudden increase on pilot workload, as well as risk of making a fault and causes pilot to lose control due to

the need of reprogramming FMS. Recommendations given based on the results of the experimental study are discussed below under the titles of design and training.

Design: FMS taking part in almost every modern cockpit nowadays is of great importance for the pilots especially in the approach and departure phases. FMS located in the lower pedestal in B-200, causes pilots' attention to focus inside in these critical phases. However, providing FMS to be positioned in the front panel will not obstruct outside scan and will not prevent interaction with flight instruments. Another significant problem to be emphasized is the usability of software systems. A usable system will result in to overcome procedures faultlessly within short times. This problem is particularly noteworthy in the critical phases of the flight.

Only one approach can be entered into FMS in the current case. In the event of a change of approach procedure, it is required to cancel previous procedure and then re-program the new one. To overcome this problem, all potential approach procedures should be defined and programmed beforehand, especially at low workload period, and can be accessed with shortcut keys and activated practically in case of a last minute change. It is therefore recommended that shortcut keys for possible approaches have to be located on FMS in standby mode.

Training: Current training programs give adequate knowledge to operate FMS. However, it can not be remarked that training is more enough to understand philosophy of automation in the situation of crisis management. To lessen these kinds of problems, it should be more focused on critical situations such as injections by utilizing simulator environment in the initial trainings.

References

1. Wickens, C.D.: Complex Systems, Process Control, and Automation" Engineering Psychology and Human Performance, 3rd edn. pp. 538–547. Prentice-Hall, Upper Saddle River, New Jersey (2000)
2. Human Performance and Limitations, JAA Airline Transport Pilot's Licence Theoretical Knowledge Manual, UK Civil Aviation Authority, London, pp. 1–6 (2001)
3. Sarter, N.B., Woods, D.D., Billings, C.E.: Automation Surprises. Handbook of Human Factors & Ergonomics, 2nd edn. pp. 1–10 (1997)
4. Billings, C.E.: Toward a Human-Centered Aircraft Automation Philosophy. The. International Journal of Aviation Psychology 1(4), 261–270 (1991)
5. Skitka, L.J., Mosier, K.L., Burdick, M.: Does Automation Bias Decision-making? International Journal of Human Computer Studies 51, 991–1006 (1999)
6. Wood, S.: Flight Crew Reliance on Automation. UK Civil Aviation Authority, 2004/10, West Sussex (2004)
7. Baberg, T.W.: Man-machine Interface in Modern Transport Systems from an Aviation Perspective. Aerospace Science & Technology 5, 495–504 (2001)
8. Sarter, N.B., Woods, D.D.: Pilot Interaction With Cockpit Automation II: An Experimental Study of Pilot's Model and Awareness of the Flight Management System. The. International Journal of Aviation Psychology 4(1), 1–28 (1994)
9. Advanced-technology Aircraft Safety Survey Report, Department of Transport and Regional Development Bureau of Air Safety Investigation, Civic Square ACT, pp. 137–216 (1999)

Combining Skin Conductance and Heart Rate Variability for Adaptive Automation During Simulated IFR Flight

Wolfram Boucsein, Andrea Haarmann, and Florian Schaefer

Physiological Psychology, University of Wuppertal, Gaußstr. 20,
42097 Wuppertal, Germany
boucsein@uni-wuppertal.de

Abstract. Adaptive automation increases the operator's workload if there are signs of hypovigilance, and takes over more responsibility in case of workload becoming too high. We refined a closed-loop adaptive system for varying the strength of turbulence in a professional simulator. In the experimental condition, twenty-four subjects flew three blocks with ten 2-min flight sections under varying turbulences. Each of the three blocks applied different combinations of autonomic measures for adaptive automation. Physiological responses were calculated every 2 min for adjusting the turbulence strength for the next 2 min, dependent on an individual setpoint. Another twenty-four yoked control subjects flew the same sequence of turbulences as the corresponding experimental subjects without adaptive automation. By combining nonspecific skin conductance responses and heart rate variability, experimental subjects' skin conductance responses oscillated very close to the individual setpoint, indicating a constant vigilance level as a result of adaptive control compared to yoked control subjects.

Keywords: Vigilance, workload, adaptive automation, human-computer interaction, psychophysiology, aviation psychology.

1 Introduction

A certain degree of the operator's attention is a prerequisite for successfully operating complex man-machine systems. Increasing the degree of automation in a system may restrict the operator's role to one of a mere observer, thus considerably reducing his/her vigilance. Furthermore, complex systems may allow for multiple modes of automation. In case of an unexpected change in situational demands or a system failure requiring immediate operator actions, he/she may not be able to perform an appropriate response since he/she may have lost situational or mode awareness. Thus, there is a need for precautionary measures to prevent an operator from vigilance decrement in case of operating automated man-machine systems.

A very powerful tool for keeping the operator's vigilance in an optimal range is adaptive automation, which refers to the capability of a system to adjust its mode or increasing/reducing the degree of automation dynamically as a consequence of changes in the operator's vigilance [1,2]. In case of hypovigilance, the system may alert the operator, thus increasing his/her attention. In turn, if the operator's workload

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 639–647, 2007.

becomes too high, the system may automatically take over more responsibility for the task in question.

Vigilance decrement as well as high workload may result in performance decrement and thus should be detectable by performance changes. However, in case of a fully automated system, no measure of the operator's performance will be available from the original task [3]. An introduction of secondary tasks would not only unnecessarily increase the operator's workload but also induce motivational problems. Since vigilance decrement is typically accompanied by a decline in psychophysiological arousal, it can be monitored by measures of central and autonomic nervous system (ANS) activity. Therefore, these measures may be used to continuously monitor the operator's attentional state. Increased workload is accompanied by an increase in arousal that is reflected in psychophysiological measures as well. Setting up an adaptive automated man-machine system will allow for both upward and downward adjustments in automation to keep the operator in an optimal state for operating the system.

Arousal shifts are reflected in various psychophysiological measures such as spontaneous electroencephalography (EEG) or cardiovascular (ECG) and electrodermal activity (EDA). Physiological recordings to be used in adaptive automation are required to be continuously monitored and on-line evaluated. They are not supposed to interfere with the task or impair the operator's well-being. First attempts to establish man-machine systems for adaptive automation in laboratory environments have used EEG derived indices [1,4,5], heart rate variability (HRV) [2,5], EDA [6,7,8], and blood pressure [9]. One important application for adaptive automation systems is conducting long-haul transport operations in an airplane. Today's commercial aircraft are flown by computer systems that allow for operating modes during which the pilots remain almost passive for long periods of flight. Such a situation is inherent to vigilance decrement, the *"out of the loop performance problem"* in system operators [10], and to *"operator hazardous states of awareness"* in aviation [4]. Because it is impractical to record EEG from commercial pilots, attempts have been made to use autonomic nervous system measures to detect vigilance decrement in pilots [11]. The goal of our current research is to provide an adaptive system, based on electrodermal and cardiovascular measures.

In a previous pilot study [12], we recorded EDA and heart rate (HR) from student subjects during four flight missions in a professional instrumental flight rule (IFR) simulator, varying the strength of turbulence in order to check the usability of ANS measures for adaptive automation. Increasing strength of turbulence resulted in an increment of nonspecific skin conductance responses (NS.SCRs) which can be interpreted as an indicator of increased workload [13]. On the other hand, progression of flight missions was associated with habituation shown by a decreased frequency of NS.SCRs and reduced sum of amplitudes. The aim of the follow-up study [14] was to construct a closed-loop adaptive system, implementing NS.SCRs as adequate arousal indicator and control variable for adjusting the strength of turbulence onset during a flight task with thirty 60-s sections. In the experimental condition, turbulences were varied according to the physiological responses of the subject, dependent on an individually predefined setpoint. Yoked control subjects received the same sequence of turbulences as their experimental counterparts, however without considering their

setpoint deviations. Paired comparisons revealed smaller deviations for the experimental subjects compared to yoked control subjects.

Our previous results look promising for the usability of autonomic measures in adaptive automation. It is, however, rather unlikely that a single physiological system will have both sensitivity and diagnosticity to cover all aspects of vigilance decrement and arousal in man-machine systems. Instead, multiple recordings from different physiological systems may be needed in order to gain a full picture of the different arousal and attentional systems [13]. In the present study, three major modifications were made to the experimental setting: Firstly, we compared different combinations of autonomic measures with respect to quality of regulation. Secondly, in order to obtain a more accurate calculation of the subjects' individual setpoint, we took four instead of two baseline recordings. The reason was that in the previous study several subjects produced more NS.SCRs during the baseline recordings under resting conditions compared to workload conditions. Using a wider range of baseline recordings was expected to take care of this problem. Thirdly, we extended recording periods from 1 to 2 min per flight section. In psychophysiological recording, very short epochs may not reliably detect changes in physiological measures.

2 Method

2.1 Subjects

Forty-eight student subjects (24 female, 24 male) aged 20-39 years (M=26.42 years, SD = 5.34 years) from different disciplines took part in the study.

2.2 Task and Design

The subjects had to accomplish the following IFR flight missions: (1) Taking off from Frankfurt/Main airport and climbing out to 2,000 feet. (2) Flying straight and level to a direction of 070, controlling altitude, speed and course. (3) After a change of altitude to 10,000 ft triggered by the instructor outside the laboratory, subjects had to turn to a direction of 060 for the final destination (Erfurt airport). (4) Keeping that course, controlling altitude and speed while facing turbulences (turbulence steps 0, 1, 3 and 5). The choice of turbulence steps was based on subjective ratings from a previous test session with 36 subjects who subjectively evaluated all six turbulence stages in counterbalanced order.

Before starting their task, subjects were familiarized with the flight simulator instruments, using no turbulences and turbulence step 3 for demonstration. Afterwards, subjects performed three blocks based on different combinations of autonomic measures in counterbalanced order: (1) NS.SCRs only, (2) NS.SCRs and HR and (3) NS.SCRs and HRV. Under (1), turbulences were modified according to deviations from an individual predefined setpoint based on NS.SCRs. Under (2), turbulence changes were triggered by deviations of NS.SCRs and HR from predefined setpoints of NS.SCRs and HR in the same direction. Under (3), deviations of NS.SCRs and HRV from their respective setpoints in opposite directions triggered the change in turbulence settings. The latter algorithm aimed at the control of artifacts that are a nagging problem in online parameterization of physiological data. In

general, NS.SCRs increase and HRV decreases under elevated arousal. Thus, a change of NS.SCRs and HRV in the same direction (e.g. a simultaneous increase or decrease in both NS.SCRs and HRV) would not trigger changes in turbulence intensity.

Prior to each block, subjects had a 2-min rest for stabilizing their physiology. Afterwards, four baseline recordings were performed (2 x 2 min without turbulences as resting period and 2 x 2 min with maximum turbulence step 5 as workload period). Next, the control computer calculated the subjects' individual setpoint for the physiological measures according to the combination of autonomic measures applied, based on the arithmetic mean of the four baseline recordings.

The subjects were divided into two groups: (1) In the experimental condition, 24 subjects flew ten 2-min flight sections per block, keeping altitude and course while facing different turbulences. Psychophysiological parameters were calculated every 2 min. They were used for triggering the strength of turbulences for the next 2 min, dependent on the setpoint of the individual subject. (2) Another 24 subjects belonged to the yoked control condition, i. e. each control subject received the same block order and sequence of turbulences as the corresponding experimental subject, regardless of his/her own setpoint and hence without adaptive automation.

The yoked pairs were always formed by either two male or two female subjects for control of gender effects. Table 1 shows the design of the study.

Table 1. Repeated measures design for different combinations of psychophysiological parameters

gender	condition	blocks (counterbalanced)		
		NS.SCRs	**NS.SCRs+HR**	**NS.SCRs+HRV**
male	experimental (adaptive)	10 flight sections, 2 min each	10 flight sections, 2 min each	10 flight sections, 2 min each
	yoked control (non-adaptive)			
female	experimental (adaptive)			
	yoked control (non-adaptive)			

2.3 Apparatus

We ran a professional IFR flight simulator software on a personal computer (LAS 5.0, made by Fahsig, Germany). The software was extended by the feature of varying the strength of turbulence by means of external control via serial port. Cockpit instruments were displayed on a 17” monitor 0.5 m in front of the subject. Controls for ailerons, elevator and throttle were provided, together with an electrical trim.

A second computer was needed for the control of adaptive automation: (1) Triggering the automatic onset and offset of turbulences on the LAS computer according to the subjects' individual setpoint (comparator function). (2) Starting physiological data recording on a third computer. (3) Receiving the on-line calculated NS.SCRs, HR and HRV from the recording computer for adaptive regulation of the subjects' arousal according to the combination of autonomic measures applied. Fig. 1 gives an overview of the information flow between the subject and the various instruments.

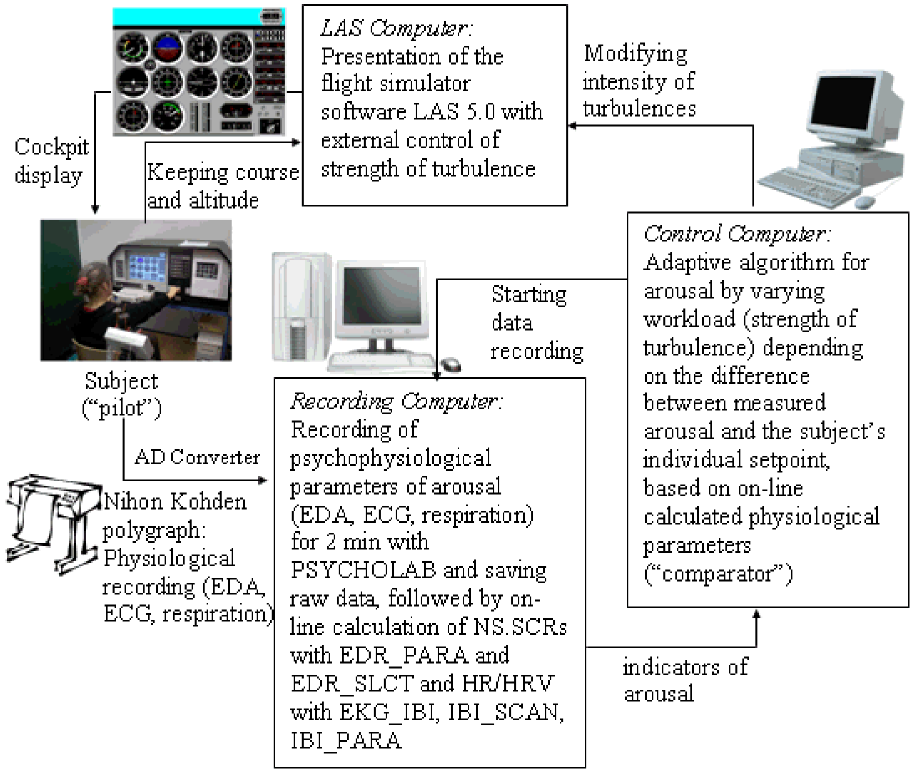


Fig. 1. Closed-loop adaptive system based on autonomic measures

2.4 Recording and Data Analysis

Recording of physiological data (EDA, ECG, respiration) was performed by means of a Nihon Kohden Neurofax EEG-8310 G polygraph, using a personal computer with a customized software package (*PSYCHOLAB*, © Jörn Grabke, 1997). EDA was recorded thenar and hypothenar according to Boucsein [15] from the left hand with two Ag/AgCl electrodes (0.8 cm diameter), using isotonic electrode cream (Med Associates, Inc.), with a sampling rate of 20 Hz, a sensitivity of 0.001 µS, and a 0.3 Hz low pass filter. Frequency and sum of amplitudes of NS.SCRs were used as tonic EDA measures, calculated on-line by customized software (*EDR_PARA* and *EDR_SLCT*, © Florian Schaefer, 2003), using an amplitude criterion of 0.01 µS. ECG was recorded by the Einthoven II-lead (above the right wrist vs. above the left ankle) with two Ag/AgCl electrodes, filled with Hellige electrode cream, at a sampling rate of 200 Hz. A ground electrode was placed on the left forearm. The ECG signal was analyzed by customized software (*EKG_IBI*, *IBI_SCAN*, *IBI_PARA*, © Florian Schaefer, 2003) calculating mean HR and HRV as root mean square of successive differences (RMSSD).

A respiration belt containing a piezo element was fastened to the subject's thorax (sampling rate of 10 Hz). Respiration was not used for adaptive automation.

2.5 Statistical Analysis

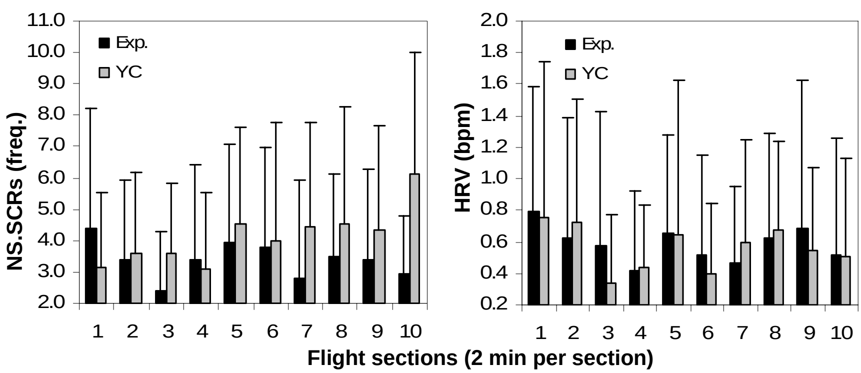
In a first step, the absolute setpoint deviation values were calculated for the psychophysiological parameters involved in the different algorithms. Afterwards, repeated measures ANOVAs were conducted separately for each block, with experimental condition (adaptive vs. yoked control), gender (male vs. female) and sequence of blocks (three combinations of physiological data) as between subject factors and the ten flight sections as within subject factor, using Greenhouse-Geisser corrected degrees of freedom.

In a second analysis, the three blocks were directly compared by repeated measures ANOVAs for each physiological measure, using only flight segments 6 to 10. That additional procedure was chosen because of marked differences between conditions within those flight sections (a posteriori) and helped to evaluate the quality of adaptive automation for the three combinations of autonomic measures applied.

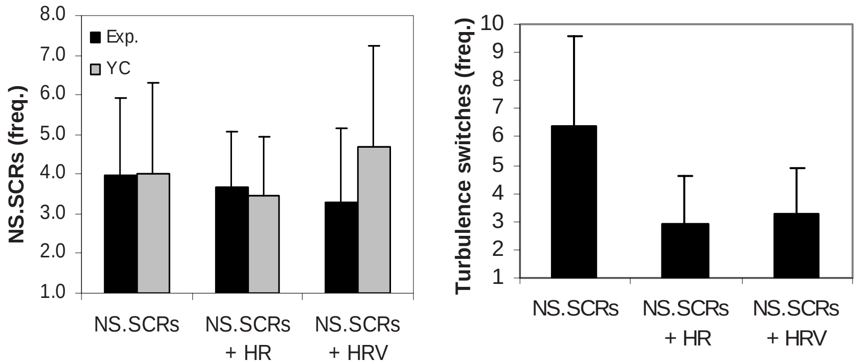
In a third analysis, the frequency of turbulence switches was calculated and submitted to another repeated measures ANOVA, with the same between subject factors mentioned above and number of switches as within subject factor. That procedure was applied to the 24 experimental subjects only as their physiological responses actually triggered the variation of turbulences in contrast to the yoked control group. The probability of error was set to α=.05.

3 Results

Repeated measures ANOVAs and post hoc comparisons revealed that during the combination of NS.SCRs and HRV, setpoint deviations of NS.SCRs were significantly smaller for the experimental group compared to the yoked control group, especially during the second half of the block (flight segments 6 to 10), as supported by a significant interaction of experimental condition and flight segment ($F(6.47, 232.72) = 2.65$, $p=.014$; $t(32.82) = 3.67$, $p=.001$; see Fig. 2). HRV data did not yield significant differences between the two groups (see Fig. 3).



Figs. 2 and 3. Absolute setpoint deviation values (group means) for adaptive (Exp.) and yoked control (YC) condition by coupling of NS.SCRs (left) and HRV (right)



Figs. 4 and 5. Comparison of the three algorithm constellations for NS.SCRs setpoint deviations (left) and frequency of turbulence switches within the experimental group (N=24) for the different couplings of psychophysiological parameters (right)

In a second analysis, a direct comparison of the three blocks using only flight segments 6 to 10 revealed a significant interaction of block and experimental condition, again with smaller setpoint deviations of NS.SCRs in experimental subjects during block "NS.SCRs+HRV" ($F(1.85, 66.75)=3.41$, $p=.042$; $t(42.15)=2.20$, $p=.033$; see Fig. 4).

An additional analysis within the adaptive automation group (N=24) revealed that the frequency of turbulence switches was significantly higher in block "NS.SCRs" compared to blocks "NS.SCRs+HR" and "NS.SCRs+HRV" with combined physiological parameters ($F(1.60, 28.71) = 17.67$, $p<.001$; see Fig. 5).

4 Discussion

The present study examined the adjustment of physiological arousal in a closed-loop system by means of different combinations of physiological measures during a simulated flight mission task in a yoked control group design. In the experimental group, adaptive adjustment was performed by means of either NS.SCRs alone, or NS.SCRs and HR, or NS.SCRs and HRV according to the subjects' individual setpoints taken from four baseline recordings. In the yoked control group, subjects flew the sequence of flight missions of their experimental counterparts without an adaptive control, i.e. regardless of their individual setpoint. Results indicated that the experimental subjects remained closer to their individual setpoint of arousal compared to yoked control subjects as a consequence of adaptive control.

The results supported the usability of autonomic measures in adaptive automation as already found in our previous study [14]. Moreover, we were able to show that coupling of two psychophysiological parameters, namely NS.SCRs and HRV, turned out to show a marked differentiation between the two experimental conditions compared to the other two constellations of algorithms with regard to setpoint deviations of NS.SCRs (see Fig. 4). Obviously, HRV had a modulating effect on

switching frequency of turbulences compared to adaptive regulation based on NS.SCRs only. In the latter case, the frequency of turbulence switches was significantly higher than under combined parameters (see Fig. 5). According to Scallen et al. [16], short cycles of automation in adaptive function allocation elevated performance, but at the same time increased subjective workload. Moreover, Hadley et al. [17] observed subjects having more difficulties in switching back from automation to manual operation in case of short cycles of switches. Hence, the high frequency of switches under NS.SCRs alone might have contributed to instabilities within the closed loop, resulting in higher deviations from setpoint values.

In addition, coupling of NS.SCRs and HRV within a single algorithm for adaptive automation can be considered as a powerful tool for counteracting artifacts during on-line assessment. Influences that are regarded as artifacts during psychophysiological recording such as body movements or deep breathing [15] will presumably cause NS.SCRs and HRV to change in the same direction. In this case, the algorithm will not initiate workload modulation. If, however, NS.SCRs and HRV change in opposite directions as a consequence of task demands, the algorithm will vary task demands according to the individual predefined setpoints. For further enhancement of closed-loop stability, the introduction of hysteresis for setpoint values should be considered.

In conclusion, our results can be considered as an important step towards the transfer of adaptive automation from the laboratory to the cockpit. We consider the advantage over the hitherto performed research, e.g. [1,4], twofold. First, we successfully used easy-to-measure autonomic variables instead of EEG measures that would be much harder to record during real flight. Second, probing adaptive automation in a professional IFR flight simulator with an authentic cockpit comes much closer to reality than using the Multi-Attribute Task Battery (MATB), an artificial system that does not provide real flight instruments, and even its tracking task does not come close to the standard-T (indicators of airspeed, attitude, altitude and direction) displays in a glass cockpit. In our opinion, more realistic setups are a prerequisite for the implementation of adaptive automation in such a complex work environment as a cockpit.

References

1. Pope, A.T., Bogart, E.H., Bartolome, D.S.: Biocybernetic System Evaluates Indices of Operator Engagement in Automated Task. Biol. Psychol. 40, 187–195 (1995)
2. Byrne, E.A., Parasuraman, R.: Psychophysiology and Adaptive Automation. Biol. Psychol. 42, 249–268 (1996)
3. Morrison, J.G., Gluckman, J.P.: Definitions and Prospective Guidelines for the Application of Adaptive Automation. In: Mouloua, M., Parasuraman, R. (eds.) Human Performance in Automated Systems: Current Research and Trends, pp. 256–263. Lawrence Erlbaum Associates, Hillsdale (1994)
4. Prinzel III, L.J., Pope, A.T., Freeman, F.G.: Application of Physiological Self-Regulation and Adaptive Task Allocation Techniques for Controlling Operator Hazardous States of Awareness. NASA Langley Research Center, Hampton, Virginia, pp. 1–17 (2001)ailable via http://techreports.larc.nasa.gov/ltrs/PDF/2001/tm/NASA-2001-tm211015.pdf

5. Prinzel III, L.J., Parasuraman, R., Freeman, F.G., Scerbo, M.W., Mikulka, P.J., Pope, A.T.: Three Experiments Examining the Use of Electroencephalogram, Event-Related Potentials, and Heart-Rate Variability for Real-Time Human-Centered Adaptive Automation Design. NASA Langley Research Center, Hampton, Virginia, pp. 1–62 (2003) available via //techreports.larc.nasa.gov/ltrs/PDF/2003/tp/NASA-2003-tp212442.pdf
6. Nishimura, C., Nagumo, J.I.: Feedback Control of the Level of Arousal Using Skin Potential Level as an Index. Ergonomics 28, 905–913 (1985)
7. Yamamoto, Y., Isshiki, H.: Instrument for Controlling Drowsiness Using Galvanic Skin Reflex. Med. Biol. Eng. Comput. 30, 562–564 (1992)
8. Isshiki, H., Yamamoto, Y.: Instrument for Monitoring Arousal Level Using Electrodermal Activity. In: Instrumentation and Measurement Technology Conference, May 10-12, 1994, IMTC/94. Conference Proceedings. 10th Anniversary. Advanced Technologies in I & M. IEEE, Hamamatsu, Japan, pp. 975–978 (1994)
9. De Waard, D., Mulder, L.J.M: Can Cardiovascular Indices be Used to Adapt the Ambulance Dispatch Worker's Task? In: Pikaar, R.N., Koningsveld, E.A.P., Settels, P.J.M. (eds.) Proceedings IEA 2006 Congress: Meeting Diversity in Ergonomics, Elsevier, Amsterdam (2006)
10. Endsley, M.R.: Automation and Situation Awareness. In: Parasuraman, R., Mouloua, M. (eds.) Automation and Human Performance: Theory and Applications, pp. 163–181. Lawrence Erlbaum, Hillsdale (1996)
11. Wright, N., McGown, A.: Vigilance on The Civil Flight Deck: Incidence of Sleepiness and Sleep During Long-Haul Flights and Associated Changes in Physiological Parameters. Ergonomics 44, 82–116 (2001)
12. Boucsein, W., Haarmann, A., Schaefer, F.: The Usability of Cardiovascular and Electrodermal Measures for Adaptive Automation During a Simulated IFR Flight Mission. Psychophysiology 42, S26 (2005)
13. Boucsein, W., Backs, R.W.: Engineering Psychophysiology as a Discipline: Historical and Theoretical Aspects. In: Backs, R.W., Boucsein, W. (eds.) Engineering Psychophysiology. Issues and Applications, pp. 3–29. Lawrence Erlbaum Associates, Mahwah (2000)
14. Haarmann, A., Schaefer, F., Boucsein, W.: Adaptive Automation Using Electrodermal Activity During a Simulated IFR Flight Mission. In: Pikaar, R.N., Koningsveld, E.A.P., Settels, P.J.M. (eds.) Proceedings IEA 2006 Congress: Meeting Diversity in Ergonomics, Elsevier, Amsterdam (2006)
15. Boucsein, W.: Electrodermal Activity. Plenum Press, New York (1992)
16. Scallen, S.F., Hancock, P.A., Duley, J.A.: Pilot Performance and Preference for Short Cycles of Automation in Adaptive Function Allocation. Appl. Ergon. 26, 397–403 (1995)
17. Hadley, G.A., Prinzel III, L.J., Freeman, F.G., Mikulka, P.J.: Behavioral, Subjective, and Psychophysiological Correlates of Various Schedules of Short-Cycle Automation. In: Scerbo, M.W., Mouloua, M. (eds.) Automation Technology and Human Performance: Current Research and Trends, pp. 139–143. Lawrence Erlbaum Associates, Mahwah (1999)

HILAS Flight Operations Research: Development of Risk/Safety Management, Process Improvement and Task Support Tools

Joan Cahill[1], Nick Mc Donald[1], Pernilla Ulfvengren [2], Franklyn Young[3], Yeray Ramos[4], and Gabriel Losa[5]

[1] Aerospace Psychology Research Group, Trinity College College Green, Dublin 2, Ireland
[2] Royal Institute of Technology, INDEK, KTH, Stockholm, S-10044 Sweden
[3] Rockwell Collins, Suttons Business Park, Reading, Berkshire, RGB, 1LA LK, England
[4] Facultad de Psicologia, Campus de Guajara S/N, La Laguna-Sta Cruz de Tenerife, 38205 Spain
[5] Iberial Airlines, PO Box E-28042, Azi Ni Barajas - Edif., Operaciones 114 Madrid, Spain
cahilljo@tcd.ie

Abstract. This paper reports on flight operations research, conducted as part of the work requirements for the Flight Operations Strand of the Human Integration into the Lifecycle of Aviation Systems (HILAS) project. Specifically, it presents a provisional framework for a suite of integrated Flight Operations tools developed in this research. It is anticipated that these tools will be used by different airline personnel to gather integrate, analyze and communicate data in relation to risk/safety management and process improvement. Further these tools will provide customized task support for different management and operational personnel.

Keywords: Flight Operations, Safety, Performance Monitoring, Risk Management, Process Improvement, Information Flow, Task Support, Human Performance.

1 Introduction

Given the continuing growth in passenger movement, there will be a doubling of air traffic within Europe by 2020 [7]. It is likely that this will lead to an increase in accident numbers. Unsurprisingly, this is unacceptable for the traveling public. Various explanations for air accidents have been propounded. Individualistic explanations focus on problems related to crew situation awareness [4], [5], task management [1], [6], crew co-ordination and communication, automation, fatigue and complacency. Systemic explanations concentrate on the different organizational factors which contribute to incidents and/or accidents. This includes commercial pressures, information failures [11], [12], poor safety culture, problems with training and process design and weaknesses in tool design. A review of specific accident reports suggests a conflux of individual and organizational factors.

Until recently, airline approaches to safety have reflected a reactive model (e.g. complying with regulatory requirements and prescribing measures to prevent the

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 648–657, 2007.

recurrence of undesirable events). Current models follow a more proactive safety management approach. According to the international civil aviation organization (ICAO), this is characterized by a number of factors including [9],

- The application of scientifically-based risk management methods
- Senior management's commitment to the management of safety
- A non-punitive environment to foster effective incident and hazard reporting
- Systems to collect, analyze, and share safety-related data arising from normal operations
- Sharing safety lessons learned and best practices through the active exchange of safety information

From ICAO's perspective, this is supported by the development of appropriate safety management systems (SMS), defining the required organizational structures, accountabilities, policies and procedures [9]. In this regard, most airlines have developed (or are in the process of developing) safety/risk management systems in accordance with regulatory guidance. This is complemented by the application of a range of system performance monitoring/evaluation tools. Currently, airlines use a range of paper and technology based tools to monitor and evaluate human performance (and by implication organizational/system safety). Feedback from these tools is used to direct system safety improvements (e.g. process/procedures re-design, enhanced training and so forth). Traditionally, these tools have divided into two types: those that focus on gathering human performance information using either self report or observer based methodologies (e.g. Air Safety Reports, Non Technical Skills Evaluation, Line Checks and Line Operations Flight Training) and those that focus gathering aircraft performance information (e.g. Flight Operations Quality Assurance). Crucially, these tools fail to provide a continuous picture of routine operations supporting predictive risk management. Further, the use of many discrete tools presents certain information management challenges. Much valuable data is gathered about the operation. Yet this data is gathered, analyzed and stored in different formats. As such, it is difficult to obtain an overall integrated safety/risk picture. Although useful from a data gathering perspective, these tools fall short of providing adequate data integration and analysis support. To this end, airlines are interested in developing tools which provide a continuous (and potentially real-time) picture of routine operations. Further, many airlines are focusing on improving knowledge integration both internally (e.g. within airline) and externally (e.g. with authorities, other airlines etc). As such, new technology concepts and information sharing practices are required to facilitate the gathering, integration, analysis and communication of all airline information (e.g. both commercial, operational and safety). From a safety management perspective, this will support the analysis of information related to ongoing management of operational risks (e.g. real time/tactical) and strategic safety/process improvement iniatives. Further, it is anticipated that this will improve airline safety culture (specifically in relation to the reporting and sharing of safety information).

In parallel, new cockpit technologies are being developed to improve flight safety. Supplemental flight information (traditionally presented in paper format and carried in the pilot's flight bag) is now being presented in digital format. This digital medium

is termed the Electronic Flight Bag (EFB). Typical EFB functionality includes electronic maps and documents, performance calculations, ground/air messaging and crew reporting. Synthetic vision technologies are being developed to enhance Flight Crew situation awareness and assist with navigation guidance and control tasks. Also, task management tool concepts, providing real-time workload assistance are being advanced [6]. Further, research is also focusing on enhancing overall cockpit information management, given the volume of information provided to Flight Crew. From a theoretical perspective situation awareness is likened to information awareness. The safety issue (and design problem) becomes one of providing crews with the right information at the right time. In this regard, timeline based information displays presenting navigational information have demonstrated safety improvements [10].

Although these tools support flight safety, they are not linked to broader airline safety/risk management tools and processes. Arguably, little or no attention has been paid to the development of cockpit task support tools which enhance (a) real-time communication between Flight Crew and other operational roles, (b) provide performance feedback (e.g. both real time and after flight) and (c) embed crew reporting in Flight Crew tasks (linking to airline safety/risk monitoring and process improvement activities). This requires redress – flight deck and airline safety practices and technologies must be integrated.

In traditional industries the need for continuous improvements in products and processes is widely recognized. Typically, this is facilitated by the development of tools and methods to increase worker satisfaction along with organisational efficiency. Airlines need to develop lean and cost-effective flight operations processes in order to increase competitiveness, while maintaining or enhancing safety and reliability. Tools and methodologies are required to acquire human factors related information from the operators, and use this information to continuously improve the process. Research suggests that the design of such tools takes second place to continuous improvement behaviour itself. This involves a suite of behaviours which evolve over time rather than a single activity [2]. These behaviours cluster around several core themes - for example; systematic finding, solving of problems, monitoring, measuring processes and strategic targeting.

2 Introduction to HILAS Flight Operations Research

2.1 Introduction to HILAS

The Human Integration into the Lifecycle of Aviation Systems (HILAS) project is part of the Sixth Framework Programme for aeronautics and space research, sponsored by the European Commission. The HILAS project will develop a model of good practice for the integration of human factors across the life-cycle of aviation systems. The project contains four parallel strands of work: the integration and management of human factors knowledge; the flight operations environment and performance; the evaluation of new flight deck technologies, and the monitoring and assessment of maintenance operations [8].

The Flight Operations strand comprises seven European airlines, human factors researchers and technology partners. The critical objective of this strand is to develop

and implement a new methodology for monitoring and evaluating overall system performance to support flight safety, operational risk management and process improvement.

2.2 Flight Operations High Level Research Plan

The flight operations research plan involves seven sequential research phases. These include:

- Phase 1: Identifying the high level requirements for the proposed system.
- Phase 2: Defining the user requirements for the proposed system.
- Phase 3: Building Version One toolset.
- Phase 4: Implementing/trialing Version One toolset with HILAS airlines.
- Phase 5: Building Version Two toolset, taking into account feedback from the different airline trials.
- Phase 6: Implementing/trialing Version Two toolset with HILAS airlines.
- Phase 7: Updating the tool framework and technologies and dissemination

Currently, phase two is nearly complete.

2.3 Overview of Phase One and Two Methodologies

The first phase of human factors research involved (a) an extensive literature review of state of the art performance monitoring and risk management tools, (b) airline process mapping (e.g. flight operations process), (c) interviews with safety and operational personnel and (d) flight operations observations. This resulted in the identification of the high level tool framework and objectives. The specific findings of phase one research have been reported in an earlier paper [3].

Phase two research involved more in depth Human Factors case studies with airline partners. HF researchers conducted extensive field research (e.g. interviews and observations) with operational and management staff involved in each of the airlines three Flight Operations sub processes. This includes flight planning, the active flight operation and the quality/improvement/safety process. In addition, technology partners conducted research in relation to airline technologies (e.g. platforms, message protocols and so forth) and tool integration requirements.

3 High Level Overview of Key Findings

This section presents a high level synthesis of field research findings with five partner airlines.

3.1 Overall Flight Operations Process Model

HILAS research shows that the Flight Operations process is structured into three related sub-processes: (a) Flight Planning, (b) Active Flight Operation and (c)

Quality/Safety/Improvement process.[1] Each sub process is divided into a series of process phases with specific critical points/states. The transition from one critical point/state (e.g. process state/aircraft state/information state) to the next requires the accomplishment of work both on an individual and team level. HILAS research indicates that Flight Operations is a complex and dynamic process subject to contingency. In this regard, the Flight Plan and associated flight context is managed by both back-office personnel (e.g. Flight Operations Control, Dispatch, Flight Planning), and front line personnel (e.g. Flight Crew, Co-ordinator, Cabin Crew, Maintenance) at different points in the flight operation. Critically, this research indicates that in certain cases, issues which arise in the active flight operation result from problems originating in the other two sub-processes (e.g. problems in the production of the flight plan, or problems handling safety or other operational feedback about specific flights/airports). This is in addition to a range of internal and external problems that can occur in real-time[2]. Crew interviews and observations suggest that some of these issues might be predicted before flight, and as such managed either before or during flight. Other issues are less predictable given the level of variability and complexity in the process. In relation to this, HILAS research suggests that there is no ‘normal’ operation, but rather a spectrum of operational and environmental complexity that constitutes normal. Further, research indicates that certain flights have higher levels of operational and environmental complexity, and for this reason may have a higher risk profile. To this end, these flights need to be managed carefully to mitigate risks.

3.2 Information Sharing/Knowledge Integration

Across the HILAS airline partners, it was noted that improved information sharing practices and technology would result in significant safety and process improvements. Overall, it was suggested that tools might facilitate both strategic and real time information sharing across the three flight operations processes. Interviews with a range of organizational functions/roles indicate that different organizational function/roles have information that is relevant to other organizational function/roles (inputs/outputs), but information sharing is not happening, or what is happening is not adequate. Participants noted that different organization functions/roles may require information in different formats, depending on the nature of their work/objectives. In relation to the production process (e.g. flight planning and active flight operations process), it seems that all roles would benefit from enhanced information sharing both from a task support and reporting perspective. Currently, it seems that information

[1] From a production perspective, the flight operations process is divided into two sequential sub processes e.g. (a) flight operation planning process and (b) the active flight operation process. A third sub process (c) the change/improvement process is linked to the other two sub processes, but is a quality/improvement process. This sub process is ongoing and runs in parallel to the other two processes. The overall flight operations process links to other processes (e.g. the aircraft turnaround/technical signoff which is part of the active flight operation links to the Line Maintenance process - part of the overall Maintenance Process).

[2] For example problems that arise due to crew errors, organization problems that arise given changes to operational context (e.g. crew changes, aircraft changes) and external problems that arise in real-time (e.g. weather, traffic, ATC restrictions).

handover is weak. Further, there is poor understanding both individually and collectively in relation to role task information requirements and constraints. In relation to safety and improvement activities, formal methods (committee meetings, report writing, email) and informal methods (private conversations, email) were identified. Nonetheless, it seems that much important information sharing occurs informally. On the whole, the possibility of sharing information between different organization functions (both for purposes of task support and process improvement) was favorably perceived. It was noted that there would be certain organizational barriers to this (e.g. attachment to the old way of doing things, preference for informal process, people being protective/territorial about information given commercial issues etc). Further, many organizational barriers in relation to performance reporting were observed. In all airlines, Pilots noted that they were unlikely to provide feedback about their own performance, unless they had to (e.g. mandatory Air Safety Reports). Pilots clearly expressed the requirement for anonymous or confidential reporting. Pilots also noted that they are not motivated to report if they do not receive feedback about the status of their reports, or if it is perceived that recommendation are not given appropriate consideration by management. Lastly, Pilots observed that the have little time for reporting. It is anticipated that issues related to confidentiality might be handled by the development of appropriate data management/protection functionality in the proposed HILAS toolset. Further, data transformation requirements might be facilitated by specific data filtering and presentation intelligence. Moreover, clever task support technologies, easily accessible to crews might reduce reporting times. Management commitment to safety and the development of a non punitive reporting culture (e.g. confidential reporting systems) is also required.

3.3 Task Support (For All Users)

Research suggests that operational and management personnel across the three flight operations process require task support (e.g. Flight Crew, Cabin Crew, Co-ordinator, Safety Department, Flight Planning, Dispatch, Flight Operations Control, Maintenance and so forth). Task support involves supporting the safe, competent, effective and timely execution of individual and collaborative work tasks/activities in relation to the achievement of the operational goal (e.g. flight planning and flight operations process) and quality/safety activities (e.g. quality/improvement process). This requires the development of tools (both technology and non technology tools), to provide information relevant to the task performance, to share information about performance, to assist in task performance and to provide feedback.

3.4 Flight Crew Task Support

Flight Crew field research shows that Flight Crew operate in a multiple task environment. Critically, they act as a coordinating interface between multiple operational roles, with different and often conflicting constraints. Although Flight Crew tasks and workflow are structured according to the logic of the operational process/timeline, actual task demands and workflow vary according to context. In this regard, Flight Crew operate at the ‘sharp end’ of the operation and require considerable task support in terms of managing operational complexity and change. HILAS research suggests that Flight Crew task performance is shaped by the quality

and availability of tools and information to hand. In this regard, Flight Crew would benefit from improved information sharing with a range of roles (e.g. Maintenance, Dispatch, Flight Planning and Flight Operations Control), both in relation to the management of operational risks, and specific collaborative task information inputs/outputs. Ideally, Flight Crew might obtain information about operational risks to be managed before flight (e.g. linked to Flight Plan). Further, Flight Crew might receive additional decision support from Operations Control, if their flight has a high risk profile, or if requested by Flight Crew. Jump seat observations of Flight Crew reveal that the quality and nature of information sharing with different operational roles varies according to social relations and context. Further, much information sharing is informal and opportunistic (e.g. Flight Crew ring Co-ordinator for flight updates using mobile phone). Flight Crew would benefit from performance feedback both in real-time and post flight. That said, this must not interfere with the primary task of flying the aircraft safely. Further, research indicates that there is limited time available for reporting. To this end, reporting should be embedded in the Flight Crew task and crews should only be required to report on safety critical events.

3.5 Process Improvement

In interviews it was clear that reporting is not an integral part of a Pilots everyday work. Yet Pilots have a lot to say about the operation and are not intimidated about speaking up. HILAS research suggests that the major barriers to reporting are lack of trust in management and insufficient time. Research suggests that reporting has the potential to be meaningful part of a Pilots job, given management commitment. All pilots agreed that they would consider reporting, if user friendly reporting tools were available during flight, and if management provided better feedback about the relationship between reporting and organisational changes. Therefore the organisation needs to:

- Regain Pilot trust (e.g. feedback from pilots is vital for process improvement).
- Handle incoming reports which facilitate learning and feedback to the operation.
- Give feedback upon receiving the report (e.g. make visible what happens with the information, how it is used and the importance of getting it).
- Clearly state who has mandate to make a change in what area.
- Give captains time during work to file reports.
- Communicate clear strategic goals and focus on improvements that are prioritised.
- Allocate resources for handling reporting and feedback process

3.6 Organizational Requirements and Culture

HILAS research suggests for the successful implementation of organisational improvement strategies, it is necessary to take into account the organisational and implementation culture. Organisational culture is manifested in shared values and meanings, and in a particular organisational structure and processes (e.g. policies, strategies, goals and practices, and leadership styles). To date, the research indicates that HILAS airlines reflect a spectrum of openness to change and innovation of rules and procedures. Significant differences were identified in relation to organisational

identity, commitment to safety, safety culture and communication strategies. This will be taken into account in the trials of the proposed toolset.

4 Overview of Tool Framework and User Requirements

To date, HILAS field research findings have been translated into a high level tool framework and specific tool user requirements. The proposed HILAS system comprises four related tools (e.g. Tools A, B, C & D) – see table 1 below. All tools link to the operational process/risk model (integrating information across the three sub processes e.g. flight planning, active operational process and change/improvement process). Tools A, B and C are technology tools. Each of these tools has a specific remit in terms of the tool objective (e.g. process improvement, risk management and task support). Tool D is a non technology tool and refers to the organizational requirements for implementing Tools A, B and C within an airline. In particular, Tool D defines the information flow for other tools.

Table 1. Description of HILAS Tools A, B, C & D

Tool	High Level Function	Users	Device
A	Human factors tools providing (1) task support, (2) performance feedback and (3) reporting capability.	Crew, Cabin crew, Maintenance, Co-ordinator	EFB, PC in Office, PDA, Mobile Phone
B	Ground server and database supporting data integration – aircraft technical data and all relevant/obtainable flight/aircraft data.	N/A	N/A
C	Data analysis and reporting tool – supporting (1) risk analysis, (2) process improvement and (3) information flow analysis.	Safety,Training, Planning,Ops,Control, Maintenance,Fuel dpt	Workstation in office
D	Organisational System. This is the overall organisational system (not a technology tool). This tool defines safety management/process improvement procedures, roles and responsibilities.	All	N/A

Tools A and C feature a range of modules. The specific functionality of each module will vary according to user role and task, user location, point in flight operations process/timeline etc. In terms of the implementation trials, airlines might choose to implement a subset of these modules. Further, individual tool modules/applications can be modified by partner airlines in line with their technology maturity and organizational environment.

HILAS airlines are at different levels of technical maturity. Further, different airlines use a range of technologies to capture data/information. In this regard, the EFB provides a platform to facilitate ground/air communications and integrate information. In order to integrate data in a common database, the server storing the

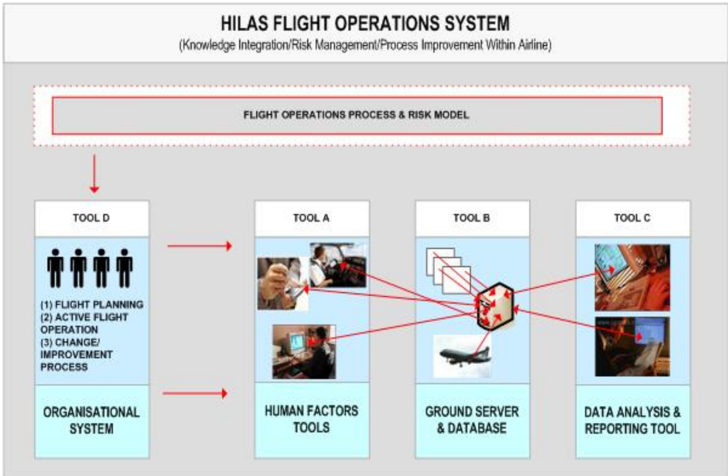


Fig. 1. HILAS Flight Operations Tools Information Flow

data will use a range of protocols to communicate with other airline systems/servers and the EFB. Data will be collected in a set of linked data records, which shall be described by a data model. The data model shall be designed to support the understanding of the system information flow at different points in the flight operations timeline – linking to the flight operations process/risk model.

5 Next Steps

The next steps involve defining the functional requirements for the proposed system (sub-set of user requirements) and developing each of the technology tools. Participatory design techniques will be used to develop the user interaction model and screen layouts for the different applications in Tools A and C. This will be conducted in parallel to application development. Also, simulations (using low and high fidelity prototypes) will be run to evaluate tool concepts. Further, additional organizational research will be conducted in relation to Tool D (e.g. organizational requirements for implementing Tools and constraints therein). Once complete, the tool-set will be trailed with partner airlines. Feedback will be elicited and a second trial conducted.

6 Conclusions

If Flight Operations is characterized by a level of risk/variability, then technology tools and organizational processes must be devised to understand the nature of these risks and identify when and how unacceptable risks/problems arise. In particular, tools must be developed to predict potential operational risks, so that safety critical events are avoided. This requires the development of knowledge integration/ information sharing tools to gather, integrate, analyze and communicate data in

relation to real time aswell as strategic safety management/oversight. It is anticipated that the suite of tools developed in this research will facilitate this.

Acknowledgments. The authors would like to thank HILAS Flight Operations Strand members and the European Commission for funding this work.

References

1. Adams, M.J., Tenney, Y.J., Pew, R.W.: Situation awareness and the cognitive management of complex systems. Human Factors 37(1) (1995)
2. Bessant, J., Caffyn, S., Gallagher, M.: An evolutionary model of continuous improvement behaviour. Technovation 21, 67–77 (2001)
3. Cahill, J., McDonald, N.: Proceedings of the International Conference on Human-Computer Interaction in Aeronautics (HCI-Aero 2006), Seattle (September 20-22, 2006)
4. Endsley, M.R, Garland, D.J (eds.): Situation Awareness Analysis and Measurement. Lawrence Erlbaum Associates, Mahwah (2000)
5. Endsley, M.R. In: Human factors in aviation operations; Proceedings of the 21st Conference of the European Association for Aviation Psychology (EAAP), Trinity College, Dublin, Ireland, 1994. Avebury Aviation, pp. 287–292 (1995)
6. Funk, K., Braune, R.: The Agenda Manager: A Knowledge-Based System to Facilitate the Management of Flight Deck Activities, 1999 World Aviation Congress, October 19-21, 1999, San Francisco, CA (1999)
7. Group of Personalities: Vision 2020 document (2001)
8. HILAS Annex 1: Description of Work. Project Number 516181. Funded by European Commission - Sixth Framework Programme (2005)
9. ICAO Safety Management Manual (Doc 9859 -AN/460) (2006)
10. Solodilova, I., Johnson, P.: Mind References in the Glass Cockpit: An Experimental Study. In: Proceedings HCI Aero 2004, Toulouse, France, ACM Press, New York (2004)
11. Turner, B.A.: Man-Made Disasters, Wykeham Publications, London (1978)
12. Turner, B.A., Pidgeon, N.F.: Man-Made Disasters. Butterworth-Heinemann, London (1997)

Analyzing Constraints to Support Computational Modeling of Air Traffic Controllers

Todd J. Callantine

San Jose State University/NASA Ames Research Center, Mail Stop 262-4,
Moffett Field, CA 94035-1000
tcallantine@mail.arc.nasa.gov

Abstract. Research toward enhancing computational models of air traffic controllers for use in assessing new air traffic management concepts is presented. Simulation data from professional air traffic controllers is analyzed using a computational method for characterizing the constraints in force when the controllers take action. The results show controllers apply strategies to proactively prevent losses of separation between aircraft. This approach has advantages for complexity reduction and workload management and implications for human roles in future air traffic management systems.

Keywords: air traffic management, cognitive work analysis, computational human performance models, constraints.

1 Introduction

The air traffic management (ATM) system is a complex socio-technical system with numerous interacting human and machine agents, including airlines, flight crews, and air traffic controllers, each using an array of technologies to support their tasks. Today's ATM system exhibits a high level of safety, but at a cost to capacity and efficiency. Major initiatives seek to increase the capacity and efficiency of the ATM system and make it more flexible and robust to disturbances (e.g. severe weather) before it is overwhelmed by growing demand [1, 2].

New ATM concepts rely heavily on advanced technologies. For example, 'trajectory-based operations' would make aircraft planned trajectories the 'currency' for information exchange throughout the system, supporting operations ranging from long-term planning to tactical separation assurance. The roles and responsibilities of humans in the system will change as new technologies are introduced. It is important not only to ensure that new operational concepts yield the desired improvements, but also that the transformed human roles—and the tools to support them—are acceptable, and provide a level of safety that meets or exceeds that of today's system.

Evaluating new operational concepts presents a challenge because the robustness and resilience of various concepts must be assessed for a range of traffic levels, aircraft equipage mixes, airspace configurations, and potential disturbances. Researchers are increasingly turning to fast-time simulations with computational agents that represent humans in the ATM system as a less expensive complement to large-scale human-in-the-loop simulations for operational concept assessment.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 658–667, 2007.

However, developing computational agents—particularly agents that produce behaviors representative of air traffic controllers—presents a challenge in itself. For example, a recent fast-time simulation demonstrated that a proposed controller strategy was workable for enabling efficient descents in heavy traffic while maintaining safety [3]. Yet in the human-in-the-loop study that followed, professional air traffic controllers opted to apply other strategies. Thus, constraints that limit the utility of the proposed strategy were omitted from the computational model.

This paper presents an analysis of constraints in force when air traffic controllers perform actions in an effort to characterize how constraints in the ATM environment shape the behavior of today's air traffic controllers. The eventual goal of the research is to better represent behavior-shaping constraints in computational agents, so that agents can be developed that support comparisons between present and future ATM concepts. Such agents may also help understand implications of automation failures and other safety-critical operations on humans in the ATM system.

Understanding constraints is a central tenet of Cognitive Work Analysis (CWA). CWA entails analyzing the work domain, control task, strategies, cooperation, and worker capabilities and limitations at multiple levels of abstraction in order to understand how constraints intrinsic to these areas shape worker behavior [4]. Effective behavior corresponds to identifying and responding intelligently to constraints and applying strategies for addressing shifting task demands. Otherwise human operators may experience high workload while exhibiting 'scrambled' control and degraded performance [5].

The paper begins by describing related research toward understanding air traffic controller behavior. The review is by no means exhaustive, but it does indicate some trends. First, most research is based on structured observations of real or simulated operations or analysis of isolated conflict situations. Second, air traffic controller conflict detection has received the most attention. Third, efforts focused on improving computational air traffic controller models are in the minority. Fourth, CWA for the ATM domain has thus far proved useful only for considering specific contextual factors and producing high-level characterizations.

By contrast, this research exercises a data-driven, computational methodology for analyzing specific behavior-shaping constraints in force when air traffic controllers perform specific actions. After describing the analysis method and presenting the results, the paper discusses how the findings inform efforts to develop improved air traffic controller agents for use in fast-time simulations of new ATM concepts.

2 Related Air Traffic Controller Research

Observations of workers performing the air traffic control (ATC) task have been conducted for more than thirty years [6]. Early research produced a description of high-level ATC functions, and provided evidence that workers in complex, dynamic environments shift the strategies they apply in response to task demands: "Less economic operating methods…progressively give way to more economic operating methods, which are not necessarily simpler, but are organized differently as work-load progressively increases. In parallel with this, the qualitative objectives that the

operators assign to themselves…are modified in a hierarchical manner, leaving priority in respect of essential objectives to the attention of the operator [6 , p. 199].”

Later research focused on developing an improved air traffic controller training curriculum, and therefore also analyzed knowledge requirements, mental models, and skills used by air traffic controllers [7]. Thirteen overlapping primary en route ATC tasks were identified and analyzed for linkages to a mental model that represents a variety of static and dynamic ATC knowledge. Sector management knowledge holds dynamic knowledge about the traffic situation and “conditions” knowledge contains factors that influence workload and strategy selection, whereas “prerequisite knowledge” contains static knowledge about airspace, procedures, and learned strategies and techniques. This research defined high-level ATC functions similarly to [6] as “to avoid (a) violation of minimum separation standards, (b) deviations from standard operating procedures, (c) disorder that may result in cognitive work overload, and (d) making unnecessary requests of the pilot [7, p. 261].” It also provided evidence that expertise at the ATC task involves the capability to work more globally, performing functions iteratively as time permitted rather than focusing solely on avoiding violations for extended periods [7].

Subsequent research has investigated how air traffic controllers construct the dynamic portion of their mental model, again with a focus on training. Heuristic methods for identifying conflicts are described in [8]. Conflict detection has also been exhaustively studied in an investigation of cognitive control levels at work during the ATC task [9].

CWA in the ATM domain has generally been limited by the difficulties inherent in describing complex dynamic constraints. It has been used to frame the current-day ATM problem at a high level and establish the importance of the trajectories an air traffic controller plans for each aircraft in guiding the planning of control interventions [10]. Other work couched in the CWA framework has examined the importance of specific sources of constraints, such as weather factors [11] and particular structural elements for different ATC operations (e.g. en route, terminal area, oceanic). Structural elements that impose constraints include standard flows, critical points, and altitude- and airspeed-based groupings of aircraft [12]. Efforts to design ecological interfaces to support airborne conflict avoidance [13] highlight some of the difficulties inherent in representing ATM constraints.

Relatively little research has been devoted specifically to improving computational models of air traffic controllers. One such effort produced a rule set to support pair-wise aircraft conflict resolution [14]. Another addressed conflict detection judgments to identify consistencies to leverage for model development [15]. Recent research also addressed the problem of air traffic controller task prioritization for the purpose of improving computational models [16]. This work considered how ‘windows of opportunity’ for performing one of six ATC subtasks could be used to identify controller task priorities. However, other environmental constraints in force at the time tasks were performed were not considered, and no clear prioritization scheme emerged from the analysis. The following section presents a computational approach for analyzing environmental constraints on air traffic controller behavior.

3 Method

A study was conducted in the Airspace Operations Laboratory (AOL) at NASA Ames Research Center to assess the viability of concept to enable efficient arrival traffic management concept under ‘rush’ conditions with significant levels of crossing traffic [17]. The airspace included three en route airspace sectors (Fig. 1). One experienced professional air traffic controller was assigned to each sector for the duration of the study, which consisted of twelve seventy-five minute traffic scenarios.

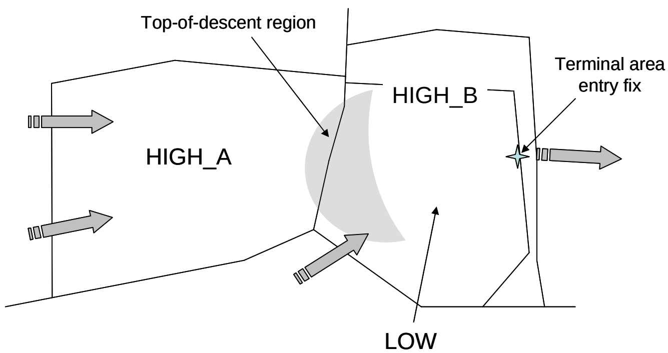


Fig. 1. Airspace used in the arrival management study. The LOW sector is situated beneath HIGH_B. Arrival flows (*arrows*) have top-of-descent points in the region shown.

Sector-specific ATC issues are illustrated in Fig. 1. The controller in HIGH_A was responsible for separating arrival aircraft from aircraft crossing the sector from the top and bottom of Fig. 1. HIGH_B issued descent clearances to arrival aircraft not already descended in HIGH_A, and also handled crossing traffic and departures climbing from the right of Fig. 1. Descending aircraft entered to the top of the LOW sector at an altitude of 24,000 feet. The LOW sector controller was responsible for merging these arrivals at the terminal area entry fix and ensuring they met speed and altitude restrictions there.

Controller workstations were high-fidelity emulations of the workstations used in current operations. Although some experimental trials included prototype decision support tools and data link communications, operations were consistent with current-day operations in all other respects; no attempt was made to retrain the controllers to control the traffic differently than they normally do.

Digital data recorded every action the controllers performed on their workstation, as well as every action pilots took in their flight simulators (other data, including voice communications, were also collected, but were not used here). The digital data were processed using a data processing tool that was developed specifically to merge and visualize data from AOL simulations [18]. The data from each experimental trial were reduced to a set of labeled events, together with the states of all aircraft when each event occurred and aircraft trajectories recorded whenever they were modified.

An enhanced version of the data processing tool used the reduced data to compute, at any point in time, deterministic trajectory predictions for each aircraft thirty

minutes into the future. The tool computed predicted trajectories at three levels of fidelity chosen to correspond with predictions air traffic controllers might make. Level 1 is a dead-reckoning prediction using the current aircraft state. Level 2 includes lateral route changes specified in the flight plan, but uses the aircraft's current ground speed and altitude. Level 3 is the highest fidelity prediction; it includes vertical information obtained using speed and altitude estimates at future points (contained in the aircraft trajectory data). The computed trajectory predictions afford a way to transform the effects of constraints to the time domain.

Constraints that impact three interdependent ATM functions—separation, flight progress, and ownership—were examined. The data processing tool computed the constraints in force, based on predictions for all aircraft in a controller's 'area of regard' (the airspace visible on the controller's workstation), for each recorded event in the data set.

Eighteen different depictions of separation constraints in force were generated for each recorded event. Nine captured the first time to a predicted loss-of-separation for each of the three levels of prediction fidelity and each of three levels of separation criteria (5, 7, and 9 nautical miles laterally, and 1000 feet vertically). The larger separation criteria were used to examine the impact of adding buffers to the legal (5 nautical mile) lateral criterion. Nine more included the effects of temporary altitudes with each of the first nine depictions. For air traffic controllers, temporary altitudes have the procedural effect of 'blocking' a set of altitude levels as 'owned' by a given aircraft (e.g. an aircraft at 33,000 feet cleared to descend to a temporary altitude of 29,000 feet is, for purposes of ensuring separation, somewhere between 33,000 and 29,000 feet). Thus, temporary altitudes may increase the possibility that controllers will perceive potential separation problems.

For a given event involving a particular aircraft, the data processing tool recorded the earliest time separation with an another aircraft in the controller's area of regard was predicted to be lost using each of the eighteen assessments. Although it is fair to assume air traffic controllers may operate with predictions at various levels of fidelity for different aircraft, trajectory predictions at different levels were not mixed in this study.

Airspace boundaries play an important role in defining ownership constraints. Typically air traffic controllers should assume control of an aircraft by the time it enters their sector. Representations of the sector boundaries were used together with the trajectory predictions to determine the time until an aircraft was predicted to enter and exit the sector.

Lastly, progress constraints were examined with respect to an aircraft's top-of-descent. Trajectory predictions with vertical information were used to assess how long before an aircraft's predicted top-of-descent an event occurred. The aircraft-computed top-of-descent point is a good estimate of latest time an aircraft can descend to a required altitude without impacting flight crew and passenger comfort. The following section presents the results of the analysis.

4 Results

After conducting a preliminary analysis on all the events for the three en route sectors, a set of action events was selected for analysis that is representative of the high-level

ATC functions and the control issues of importance in the sectors of interest. Table 1 describes the selected events.

Table 1. Action events selected for analysis. Marked (*) events are pilot actions which strongly indicate, but do not guarantee, a controller issued a corresponding clearance.

Event Label	Description
SET_ALT	Set altitude: Controller issues a clearance to the given altitude
TEMP_ALT	Temporary altitude: Controller issues a temporary altitude, blocking altitudes between the current altitude and the temporary altitude
FMS_CRZ_SPD*	Program new cruise speed: Suggests controller has issued a speed clearance to an aircraft at cruise altitude
FMS_DES_SPD*	Program new descent speed: Suggests controller has issued a speed clearance to an aircraft about to descend, or already descending
SPD_SEL*	Speed select: Suggests controller has issued a speed clearance
RTE_AMEND	Route amendment: Controller issues a new route clearance to an aircraft (typically direct to a down-path waypoint)
TGT_HDG*	Target heading: Suggests controller has issued a heading vector
TP_START	Start trial planning: Controller displays a ‘trial plan’ for an aircraft, either to see where it plans to go, or to modify its trajectory
TP_INSRT_PT	Insert point in trial plan: Controller plans to alter the aircraft’s trajectory
TP_UPLNK	Uplink trial plan: Controller data-links a new trajectory to an aircraft
HO_INIT	Initiate handoff: Controller initiates coordination with downstream controller in preparation for transferring control of an aircraft
HO_ACCPT	Accept handoff: Controller notifies upstream controller of readiness to accept control of an aircraft

The events in Table 1 are grouped by functions they support and the control method they correspond to. SET_ALT and TEMP_ALT support separation and (for arriving aircraft) flight progress and correspond to altitude control. FMS_CRZ_SPD, FMS_DES_SPD, and SPD_SEL primarily support separation (and, at a lower level, arrival sequencing and spacing) and correspond to speed control. RTE_AMEND and TGT_HDG correspond to strategic and tactical lateral control, respectively. RTE_AMEND supports both flight progress and separation, while TGT_HDG supports separation. The ‘TP_’ events signify route amendments, but may also be used for refining trajectory predictions, as well as ensuring separation. Lastly, HO_INIT and HO_ACCPT support the ownership function (a prerequisite to issuing clearances for separation and flight progress).

Fig. 2 shows the numbers of the selected action events performed in each of the three sectors considered in the analysis. Trial planning (‘TP_’) actions took place

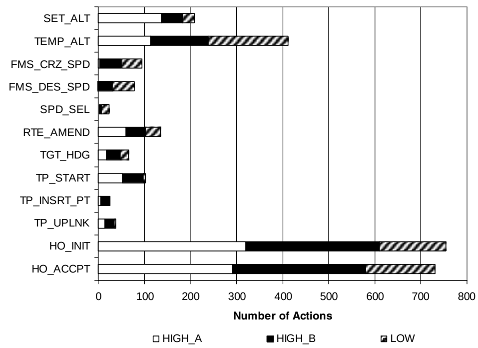


Fig. 2. Number of selected actions of each type recorded for air traffic controllers in the three en route sectors

primarily in the high altitude sectors, while speed clearances were issued primarily in HIGH_B and LOW. Altitude clearances, route amendments, and ownership actions are well represented in the data for all three sectors.

Because air traffic controller behavior exhibits considerable variability due to differences in operational context and individual performance, measures of central tendency were computed. Box plots depict values for the minimum, first quartile, median (*diamond*), third quartile, and maximum values for each of the measures below.

Fig. 3 shows the predicted times to loss-of-separation for Level 3 predictions and a seven nautical mile lateral separation for each of the three en route sectors. Controllers performed some actions in the last five minutes before a predicted loss of separation; however, for the most part controllers took action with no losses of separation predicted for the entire thirty minute horizon. These results also indicate that although some sector-specific differences are evident, median loss-of-separation times are consistently large. Therefore, additional results will be considered in the aggregate for all three sectors.

Fig. 4 shows the predicted times to loss of separation for three prediction schemes, including the most likely to predict impending conflicts (Level 3 with temporary altitudes considered, and nine nautical mile lateral separation). Even such conservative predictions show median times to potential losses of separation lie more than fifteen minutes in the future. As expected, more tactical conflict resolution activities (i.e. TGT_HDG, TP_INSRT_PT) show a lower first quartile value than other actions.

Fig. 5 shows results for constraints related to progress and ownership. These results again show a tendency toward taking action early. For example, controllers performed more than one fourth of TEMP_ALT and RTE_AMEND actions immediately after gaining control of aircraft, before the aircraft entered the sector. The next section discusses the implications of these results.

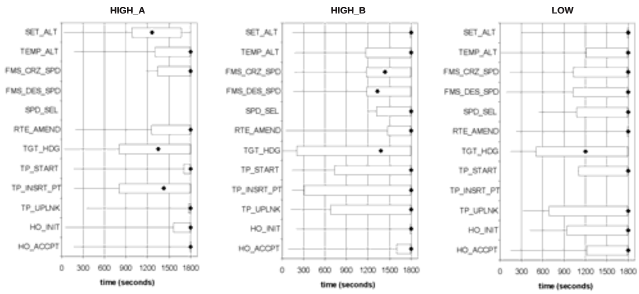


Fig. 3. Predicted times to loss-of-separation for each sector using Level 3 predictions and a 7 nautical mile lateral separation criterion

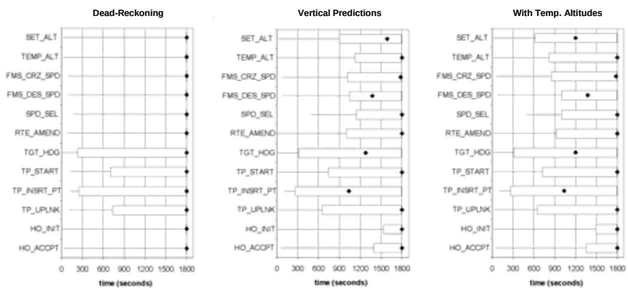


Fig. 4. Aggregate predicted times to loss-of-separation using a 9 nautical mile lateral separation criterion for Level 1 (*Dead-Reckoning*), Level 3 (*Vertical Predictions*), and Level 3 with temporary altitude predictions (*With Temp. Altitudes*)

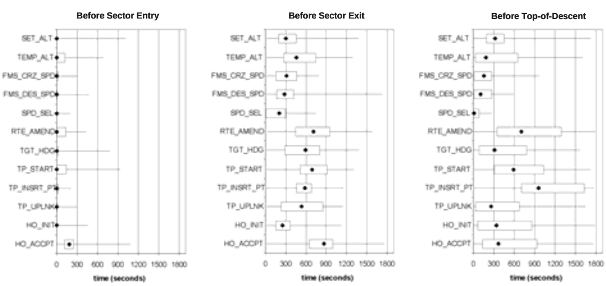


Fig. 5. Aggregate predicted times before aircraft enters the sector (*Before Sector Entry*), leaves the sector (*Before Sector Exit*), and needs to start descending (*Before Top-of-Descent*), all using Level 3 predictions

5 Discussion

The results indicate that controllers apply strategies that diminish the criticality of conflict detection and resolution. These strategies might be termed ‘confliction

prevention' strategies. Upon first attending to an aircraft, controllers attempt to 'look all the way through the sector' in order to identify a clearance that will enable the aircraft to traverse the sector without further intervention. The more often controllers are able to do this successfully, the more time they allow themselves to address isolated problems when they arise, and the less complex the problems are, because some aircraft are already effectively removed from consideration. In addition, it helps ensure that downstream controllers also have the opportunity to address problems early.

A process model of current day air traffic controller performance should therefore include this conflict prevention step. The proposed strategies modeled in [3] enabled efficient descents, but constituted a 'wait-and-see' approach to addressing potential conflicts that the results of this research indicate controllers are not comfortable with.

The results also have implications for the development of future ATM concepts and controller tools to support them. First, controllers are likely to embrace tools that support trajectory prediction in support of conflict prevention. Second, controllers are likely to have difficulties if operations force them to abandon conflict prevention, unless trustworthy automation tools support 'just-in-time' operations. Both of these assessments are in line with prior research on automation tools for air traffic controllers [19].

The results also suggest that a similar computational approach may be applicable to identifying high-workload conditions in which controllers lack the time to apply conflict prevention strategies to aircraft entering their sector. The approach may also be useful for elucidating the structure of internal planning and workload management constraints to support computational modeling.

6 Conclusion

The paper presents an analysis of constraints in force when professional air traffic controllers take action, using data from a recent NASA ATM simulation. A data-processing tool enhanced with trajectory prediction capabilities is used to reconstruct constraints related to ensuring separation, managing flight progress, and assuming ownership of aircraft. The results indicate that current operating practice includes a conflict prevention component helpful in managing workload.

Acknowledgments. This research was supported by the Super Density Operations Project of the NASA Airspace Systems Program. Thanks to Dr. Everett Palmer and Dr. Paul Lee at NASA Ames Research Center for useful discussions contributing to this research.

References

1. Joint Planning and Development Office: Concept of Operations for the Next Generation Air Transportation System. Draft Version 0.2 (July 24, 2006)
2. SESAR Consortium: SESAR Definition Phase Deliverable 2, Air Transport Framework, The Performance Target. SESAR Document DLM-0607-001-02-00 (December 2006)
3. Callantine, T., Palmer, E., Homola, J., Mercer, J., Prevot, T.: Agent-Based Assessment of Trajectory-Oriented Operations with Limited Delegation. In: Proc. 25th Digital Avionics Systems Conf., Portland, OR (2006)

4. Vincente, K.J.: Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work. Lawrence Erlbaum Associates, Mahwah (1999)
5. Hollnagel, E.: Human Reliability Analysis: Context and Control. Academic Press, London (1993)
6. Sperandio, J-C.: The Regulation of Working Methods as a Function of Work-load among Air Traffic Controllers. Ergonomics 21(3), 195–202 (1978)
7. Seamster, T.L., Redding, R.E., Cannon, J.R., Ryder, J.M., Purcell, J.A.: Cognitive Task Analysis of Expertise in Air Traffic Control. Int. J. Aviation Psychology 3(4), 257–283 (1993)
8. Nunes, A., Mogford, R.H.: Identifying Controller Strategies that Support the Picture. In: Proc. 47th Annual Meeting of the Human Factors Society, pp. 71–75 (2003)
9. Morineau, T., Hoc, J-M., Denecker, P.: Cognitive Control Levels in Air Traffic Radar Controller Activity. Int. J. Aviation Psychology 13(2), 107–130 (2003)
10. Dowell, J.: Formulating the Cognitive Design Problem of Air Traffic Management. Int. J. Human-Computer Studies 49, 743–766 (1998)
11. Ahlstrom, U.: TRACON Controller Weather Information Needs: II. Cognitive Work Analysis. DOT/FAA/CT-TN04/09. FAA William J. Hughes Technical Center, Atlantic City International Airport, NJ (2004)
12. Davison, H., Histon, J., Ragnarsdottir, M.D., Major, L., Hansman, R.J.: Impact of Operating Context on the Use of Structure in Air Traffic Controller Cognitive Processes. In: Proc. 5th U.S.A./Europe Air Traffic Management R&D Seminar. Budapest, Hungary (2003)
13. Van Dam, S., Abeloos, A., Mulder, M., van Paassen, M.: Functional Presentation of Travel Opportunities in Flexible Use Airspace: An EID of an Airborne Conflict Support Tool. In: Proc. IEEE Intl. Conf. on Systems, Man, and Cybernetics. The Hague, Netherlands, pp. 802–808 (2004)
14. Mondoloni, S.: Development of an Enroute Conflict Resolution Rulebase for the Reorganised Air Traffic Control Mathematical Simulator. NARIM-A09008-01, CSSI, Inc., Washington, DC (1998)
15. Averty, P.: Conflict Perception by ATCS Admits Doubt but not Inconsistency. In: Proc. 6th U.S.A./Europe Air Traffic Management R&D Seminar. Baltimore, MD (2005)
16. Rantanen, E.M., Levinthal, B.R.: Effect of Air Traffic Controller Taskload and Temporal Awareness on Task Prioritization. In: Proc. Intl. Symposium on Aviation Psychology, pp. 601–606 (2005)
17. Prevot, T., Callantine, T., Homola, J., Lee, P., Mercer, J., Palmer, E., Smith, N.: Air/Ground Simulation of Trajectory-Oriented Operations with Limited Delegation. In: Proc. 7th U.S.A./Europe Air Traffic Management R&D Seminar. Barcelona, Spain, (Submitted) (2007)
18. Callantine, T.: Air Traffic Management Simulation Data Visualization and Processing Tool. AIAA-2005-6489. American Institute of Aeronautics and Astronautics, Reston, VA (2005)
19. Prevot, T.: On the Design of Integrated Air/Ground Automation. In: Proc. IEEE Intl. Conf. on Systems, Man, and Cybernetics. Kona, HI, pp. 3933–3938 (2005)

Risk-Based Information Integration for Ship Navigation

Boris Gauss and Matthias Rötting

Berlin University of Technology, Department of Human-Machine Systems,
Sekr. 2-7/1, Franklinstr. 28-29, 10587 Berlin, Germany
{gauss, roetting}@mms.tu-berlin.de

Abstract. The Navigational Risk Detection and Assessment System (NARIDAS) is an approach to risk-based information integration on the ship's bridge. The purpose of this novel system is to reduce data overload and to support situation awareness of the bridge team. This paper focuses on the evaluation of NARIDAS during the development process. Evaluation is performed with system prototypes and practitioners. Three levels of evaluation are addressed: risk model validity, graphical user interface (GUI) design, and system usability. In two evaluation studies, positive results were obtained on all three levels. These results suggest that NARIDAS provides a valid model for the risks of ship navigation, a well-designed GUI, and a high usability for enhancing situational risk awareness of the bridge team.

Keywords: Ship Navigation, Risk, Support System, Evaluation, Situation Awareness, Information Integration.

1 Introduction

Approximately 80% of maritime accidents are attributed to "human error". Analyses show that many of these accidents occurred because the bridge team had lost situation awareness [1]. Situation awareness can be defined as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [2]. We assume that an important reason for loss of situation awareness on the ship's bridge is data overload, caused by ill-designed human-computer interaction. Data overload is considered a significant problem in many domains of human-computer interaction [3]. Modern ships are equipped with extensive technological aids for navigation (e.g., Automatic Radar Plotting Aids, Electronic Chart Display and Information System, Automatic Identification System). Due to the prevailing technology-centred approach to system development, usability of these systems is often low. Practitioners find the design of modern ships bridges an "ergonomic nightmare" [4].

In consequence, support provided by existing navigation aids is limited. If we consider the function of existing systems for the information processing of the human operators, we find an ever-growing number of 'information acquisition systems' on the bridge (Stage 1 of human information processing, Fig. 1). More and more information is acquired by technological systems, but the tasks to assess the information from multiple sources (i.e., to achieve and maintain situation awareness),

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 668–677, 2007.

and to decide what to do next remain with the human operators. The bridge team cannot profit any more from the very fast and accurate numerical description of the ship navigation process, because the overabundance of data presented by technological systems exceeds their cognitive capacity.

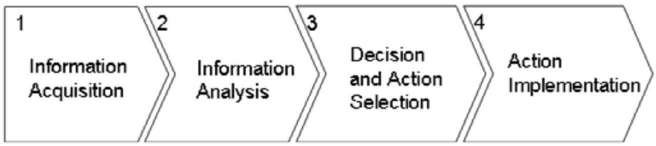


Fig. 1. Four stages of human information processing [5]

A possible solution to the problem of data overload, caused by too many "information acquisition systems", is the development of support systems for the cognitive processing stage of information analysis. On this stage, the information acquired on the first stage is integrated by relating it to the current goals. The operators extract the *meaning* of the information in their task environment for decision and action selection. An important semantic category at this stage is the concept of *risk*. Risk can be defined as the anticipation of an event with negative consequences. In dynamic human-machine systems, subjective risk assessments are directly linked to decision making and action. If subjective risk is too high, the operator will change his or her plan and take adjusting actions to reduce risk to an acceptable level [6]. Of course, it is crucial for adequate decision making that risk is assessed correctly, i.e., that the operators' subjective risk reflects the situation's actual or "objective" risk. Thus, an "objective" risk assessment system could support the cognitive processing stage of information analysis in order to overcome the data overload problem. In addition to the raw sensor data of information acquisition systems, a risk assessment system offers a task-oriented integration of the acquired information.

The rest of this paper is organised as follows: In section 2, we provide a short description of the Navigational Risk Detection and Assessment System (NARIDAS). Section 3 is dedicated to the evaluation framework within the NARIDAS development process. Section 4 presents the procedure and the results of two evaluation studies. In section 5, we discuss our results and outline some perspectives for further research and application.

2 The Navigational Risk Detection and Assessment System

For ship navigation, the *Navigational Risk Detection and Assessment System* (NARIDAS) is an approach to support integration of nautical data by dynamic risk assessments. The basis of NARIDAS is the breakdown of the bridge team's navigation task into eight task dimensions [7]:

- COLLISION AVOIDANCE (COL): pass other ships or objects safely
- ANTI-GROUNDING (GRD): adjust own ship's speed to the natural conditions

– TRACK KEEPING (TRA): keep track and consider manoeuvering area
– TRAFFIC (TRF): account for characteristics and density of traffic
– BRIDGE MANNING (MAN): consider the condition of the bridge crew
– ENVIRONMENT (ENV): account for the meteorological and hydrological conditions
– ENGINE/WHEEL (ENG): consider the state of propulsion and rudder engines
– ECONOMY (ECO): comply with the economic criteria of the voyage

For each of these task dimensions, NARIDAS calculates the corresponding risk by means of knowledge-based and rule-based procedures. In a first step, approximately 100 technical or physical input parameters – that are continuously updated from various sources (e.g., radar, electronic chart, integrated navigation system) – are processed by crisp mathematical algorithms for nautical calculations. In doing so, the input parameters are integrated into 24 higher-order variables. These higher-order variables are further processed with fuzzy algorithms comparing their current values with standard values for "good seamanship" to obtain the eight navigational risk values on a scale from 0="No Risk" to 1="Accident".

Fig. 2. NARIDAS graphical user interface

On the graphical user interface (GUI) of NARIDAS, the eight situational risk values are displayed in a bar graph (Fig. 2). This comprehensive display allows for an assessment of the situational risks of the navigation process at a glance. In addition, the system offers access to more detailed explanations, so the users can check the reasons behind the system's risk assessments.

Since navigational risks are context-specific, the NARIDAS knowledge-base is customised on three different levels: (1) long term: to the particular ship (manoeuvring properties, engine characteristics etc.), (2) medium term: to the voyage plan (way points, estimated time of arrival etc.), and (3) short term: to the current sea area. For the latter, NARIDAS distinguishes between six different "navigation modes" ("coastal waters", "approach", "traffic separation scheme", "fairway", "open sea", "at anchor"). For each navigation mode, a specific knowledge base is activated.

3 Evaluation Framework

To avoid the problems resulting from technology-centred development ("ergonomic nightmare", see above), NARIDAS is developed in a parallel-iterative process. From early stages of the development process, we work on technological and human aspects of the system in parallel. The match of these aspects is controlled in iterative evaluation loops with prototypes and the participation of practitioners. The main objective of the evaluation is formative, i.e., to gather information about how to improve the system. Evaluation in the NARIDAS development process can be assigned to an 'evaluation pyramid' of three levels (Fig. 3). On the basic level, the validity of the NARIDAS risk model is verified. Secondly, the design of the graphical user interface (GUI) is reviewed. Finally, the usability of the complete system is evaluated.

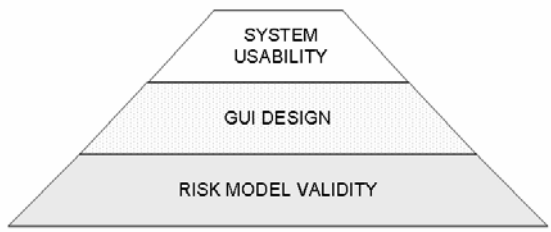


Fig. 3. Evaluation pyramid

3.1 Risk Model Validity

NARIDAS was invented by a domain expert, Dr. D. Kersandt, on the basis of his vast experiences on board ships as Master and Nautical Officer as well as in the academy as nautical instructor and accident researcher [7]. He designed and adjusted the NARIDAS knowledge base, i.e., the algorithms for risk calculation, in several years of development work. The evaluation objective at the basic level is to check how well the algorithms represent the risk assessments of other nautical experts. Of course, it is an essential prerequisite for the usability of NARIDAS that the risk algorithms reflect the common view on risk, and not just the personal opinion of a single expert. Key criteria on this level are sensitivity and selectivity of the model's risk assessments. Sensitivity refers to the degree to which the model distinguishes between different states of risk, in particular, the degree to which it detects states of "objective" high risk. Selectivity is the degree to which the model is sensitive only to changes in "objective" risk.

3.2 Graphical User Interface Design

The most important question at this level is how the risk values should be displayed to provide an optimal overview of the situation. Also, the presentation of the additional information (e.g., details of risk calculations, explanation components), the menu structure, and the GUI's conformity with general dialogue principles according to ISO

9241-111 [8] (e.g., controllability, error tolerance, suitability for learning) have to be evaluated.

3.3 System Usability

Usability is defined as the “extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [9]. The two basic levels of the evaluation pyramid can be considered necessary conditions for system usability. But a valid risk model and a usable GUI are not sufficient for usability of the complete system. To evaluate usability, the three criteria of effectiveness, efficiency and satisfaction have to be specified and measured. While satisfaction can only be assessed with “soft” subjective judgments on the system by its users, effectiveness and efficiency should be confirmed by “hard facts” where possible. However, effectiveness and efficiency of a risk assessment system are not easy to prove. In particular, the economic benefits (i.e., efficiency) are hard to calculate prospectively. As a first step, we focus on the effectiveness of the system in terms of its effects on situational risk awareness and navigation performance of the bridge team.

4 Evaluation Procedure and Results

Until now, two studies were conducted in the NARIDAS evaluation process. Study I addressed the two basic levels of the evaluation pyramid. Study II investigated the top level.

4.1 Study I: Evaluation of Risk Model and GUI Design

Study I was carried out in two rounds with a functional NARIDAS prototype, which presented the GUI, and contained the nautical data and risk values for several pre-defined static traffic scenes. The objectives of the study were (1) to compare the NARIDAS risk values with risk judgments of nautical experts and (2) to enquire the experts’ opinions about the GUI and the overall concept of this nautical risk assessment system.

Procedure. Participants were 16 nautical experts (masters, mates, final-year students) of German nationality. All of them were between 25 and 60 year-old men with nautical experience on board of large vessels world-wide. The study was conducted in individual trials. After an introduction to NARIDAS, 14 static traffic scenes were presented to the expert. These scenes represented a broad range of different navigational requirements (e.g. passing Straight of Gibraltar; approaching port of Livorno; open sea) and environmental conditions. For each scene, the experts received data about own ship characteristics (pilot card), traffic situation and sea area (screenshots of radar and electronic chart), and environmental data (wind, waves, visibility etc). Experts were instructed to judge the navigational risks of the traffic scene on the eight dimensions. After the risk assessment, a computer screen with the functional prototype was switched on, so that the experts could explore the system and compare their own risk assessments with the NARIDAS values. During risk

assessment and system exploration, experts were asked to think aloud. Verbal data was recorded, transcribed and analyzed qualitatively. After completion of the risk judgments, a detailed usability questionnaire with rating questions (5-points Likert-Scale) was administered. At the end of each trial a short structured interview was held on the expert's opinion about NARIDAS.

Results. Over all traffic scenes and risk dimensions, experts' judgments and NARIDAS values were highly consistent (Cronbach's Alpha between .89 and .94). For analysis of sensitivity and selectivity, rates of "*misses*" (sensitivity) and "*false alarms*" (selectivity) were determined.

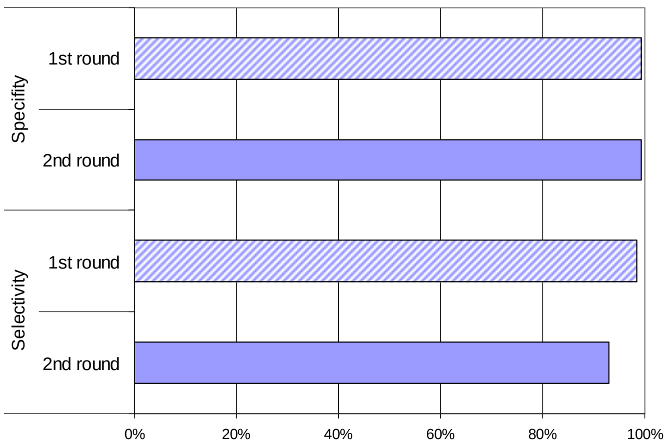


Fig. 4. Sensitivity and selectivity in both rounds of study I

A *miss* was defined as a case if >50% of the experts assessed a risk as "dangerous" (>.80) and NARIDAS assessed the risk as "not dangerous" (<.60). A false alarm was defined as a case if NARIDAS assessed a risk as "dangerous" and >50% of the experts assessed the risk as "not dangerous". With 0.9% of misses (both rounds), and 1.5% (1^{st} round) and 7.1% (2^{nd} round) of false alarms for a total number of 112 cases (14 scenes*8 risk values), sensitivity and selectivity of NARIDAS were high (Fig. 4).

In the questionnaire, the GUI was rated very positive. Participants judged the NARIDAS interface as clearly designed and easy to use. Overall usability of the system, assessed on a 10-items-scale (e.g., "NARIDAS is a reliable system", "NARIDAS would enhance the safety of navigation"), achieved 73.2 of 100 points. Also the qualitative data (think-aloud protocols, interviews) showed that the experts considered NARIDAS as a useful support to ship navigation.

4.2 Study II: Experimental Evaluation of System Usability

For study II, a highly-developed NARIDAS prototype was implemented in the full-mission ship-handling simulator in Elsfleth (Lower Saxony, Germany) (Fig. 5). The Elsfleth simulator provides four interconnected, fully equipped ship's bridges, two of

Fig. 5. NARIDAS in the Elsfleth Simulator

them with visual simulation system. These two were used for the present study. The objectives of the study were to investigate the effects of NARIDAS on situational risk awareness and navigation performance of the bridge team during a simulated voyage.

Procedure. NARIDAS was connected to the simulator network, so it was calculating the risks online during the whole voyage. Participants were 23 nautical students in the final year of their studies (all men; age between 21 and 48 years, mean=28 years). They were grouped into 11 bridge teams, each team consisting of one “Master” and one or two “Watch Officers”. A traffic scenario in the English Channel of 80 minutes was constructed with high traffic density and rather unpleasant environmental conditions (4m swell from 220°, 30kn wind from 180°, 2.5kn current from 50°). Own ship was a container vessel traveling from Cadiz to Rotterdam.

A simple one-factor experimental design was realized with “NARIDAS support” as independent variable, which was varied within teams. Each team traveled one 40-minutes section of the trip with NARIDAS, the other 40-minutes section without NARIDAS. The sequence of sections with and without NARIDAS was balanced between teams. Dependent variables were assessed with a combination of different methods. After each section, rating questionnaires were applied to assess situational risk awareness (SRA) and navigation performance (self-ratings by the subjects, and assessment of the teams by an experienced instructor). Furthermore, SRA was measured with an online-test, 3 times during each voyage section (after 15, 25 and 35 minutes). For this test, the ‘Master’ received a phone call from the experimenter. He was asked to report the three most dangerous risks at the particular moment, and to rate these risks on a scale from 0 to 100. The answers were recorded, and categorized ex post to the NARIDAS risk dimensions for analysis. As an additional indicator for navigation performance, NARIDAS risk values were recorded during the whole trip.

Results. In the SRA online-test, subjects had higher risk awareness in the sections traveled with NARIDAS support (Figure 6). In particular, more collision risks (i.e., dangerous radar targets) were reported by the participants. The difference between the

sections with and without NARIDAS is statistically significant (Wilcoxon-Test, $p<.01$). Results also show that only three of the eight risk dimensions (collision, environment, and traffic) were rated "dangerous" during the test. This indicates that overall complexity of the traffic scenario was rather low. The teams were able to handle the requirements of this simulator exercise without major problems.

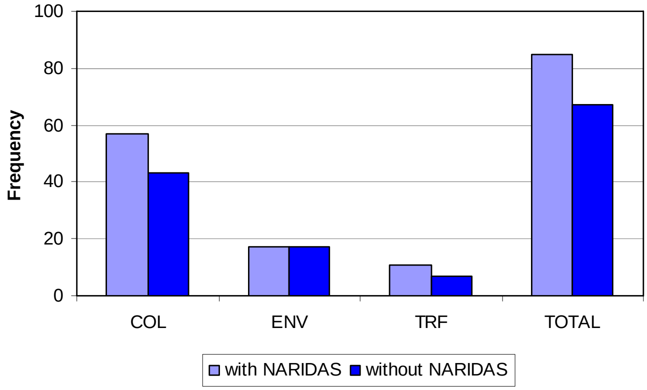


Fig. 6. Reported risks in the SRA online-test

Analysis of navigation performance showed that with NARIDAS, a higher risk of collision (the most important risk dimension in the scenario) was associated with better SRA and navigation performance ratings by the instructor. Without NARIDAS, a higher risk of collision was associated with a poorer instructor rating (Table 1). This result suggests that NARIDAS can contribute to a better handling of high risks. If a high risk is taken consciously (i.e., with a high SRA, supported by a risk assessment system), navigation performance is good, and the situation remains under control. In contrast, if the navigators take a high risk without recognizing it (lower SRA, no support), navigation performance becomes unstable.

Table 1. Correlations between instructor ratings and recorded COLLISION risks

Correlations (Spearman-Rho, *p<.05)		COLLISION risk With NARIDAS	Without NARIDAS
Instructor Ratings	Navigation Performance	.48	-.59*
	Situational Risk Awareness	.68*	-.45

Usability of NARIDAS was rated positive by the participants, and their satisfaction with the system was high. In an overall judgment, 19 participants rated NARIDAS as "good" or "very good", the other 4 participants as "neither good nor bad". There were no negative judgments on this novel system.

5 Discussion

In the two empirical studies, results were encouraging on all three levels of evaluation. In study I, the risk values calculated by NARIDAS matched very well with the risk judgments of nautical experts. Sensitivity and selectivity were high. These results indicate that the NARIDAS risk model is valid. Furthermore, these findings imply that there is a common view on the navigational risks among nautical experts, and this common view can be modeled by a combination of mathematical and fuzzy-set algorithms. However, it should also be noted that consistency of risk assessments between the experts and NARIDAS, as well as inter-individual consistency between the different experts, is high but not perfect. If we use more abstract concepts like risk, we will be confronted with a higher degree of uncertainty than with crisp technical or physical parameters (e.g., ship's speed, course, position). In complex, dynamic processes like ship navigation, human operators will always have to cope with uncertainty. The concept of risk makes uncertainty measurable and visible. The positive expert ratings on user satisfaction in both studies suggest that practitioners believe they will profit from the display of risks by NARIDAS, despite the residual fuzziness of the risk concept.

For study II, NARIDAS was successfully implemented in the full-mission ship-handling simulator Elsfleth, so the system's operational capability could be demonstrated online in a dynamic setting. Experimental comparison showed positive effects of NARIDAS on situational risk awareness and navigation performance, even though the voyage scenario realized for the simulator study resulted to be not extraordinarily challenging for the well-trained participants. In the future, NARIDAS effectiveness should be tested under more tricky conditions, e.g. a slowly evolving emergency scenario in a simulator exercise of several hours. We assume that the benefits of NARIDAS should appear even clearer if the bridge team had to switch unexpectedly from operational routine to a peak workload situation. In study II, workload was rather moderate without major variations during the exercise, reflecting an everyday's working scenario.

Furthermore, the measurement of navigational risks by NARIDAS offers perspectives for various applications beyond the use as support tool for the bridge. In the ship-handling simulator, NARIDAS could provide online training feedback for the students as well as standardized assessments of navigation performance. Last but not least, dynamic risk assessments could be integrated into voyage data replay systems. So, incident and accident analyses would profit from risk profiles of critical situations, e.g. for a quantitative determination of the 'point of no return'.

References

1. Baker, C.C., McCafferty, D.B.: Accident Database Review of Human-Element Concerns: What Do the Results Mean for Classification? In: Proceedings of Conference on Human Factors in Ship Design, Safety & Operation. RINA, London, pp. 65–71 (2005)
2. Endsley, M.R.: Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors 37(1), 32–64 (1995)

3. Woods, D.D., Patterson, E.S., Roth, M.: Can We Ever Escape from Data Overload? A Cognitive Systems Diagnosis. Cognition, Technology and Work 4, 22–36 (2002)
4. The Nautical Institute: Alert! The International Maritime Human Element Bulletin 3 (April 2004) [visited 6 February, 2007] http://www.he-alert.org/
5. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A Model for Types and Levels of Human Interaction with Automation. IEEE Transactions on Systems, Man, and Cybernetics – Part. A: Systems and Humans 30(3), 286–297 (2000)
6. Näätänen, R., Summala, H.: Road-User Behavior and Traffic Accidents. North-Holland, Amsterdam (1976)
7. Kersandt, D.: ANRIS – Automated Navigation Risk Identification System. A Knowledge-based Early Warning System for Avoiding the Point of No Return. In: The Influence of the Man-Machine-Interface on Safety of Navigation. Proceedings of ISHFOB ’95 International Symposium of Human Factors on Board (1995)
8. Ergonomics of human-system interaction – Part 110: Dialogue principles
9. Ergonomic requirements for office work with visual display terminals – Part 11: Guidance on usability

Experimental Thermal/Moisture Mapping of Industrial Safety Helmets

Z.W. Guan, A.R. Dullah, and H.L. Zhou

Department of Engineering, University of Liverpool, U.K
Zhongwei.Guan@liv.ac.uk

Abstract. This paper presents the research on thermal/moisture mapping of typical industrial safety helmets using match-head-sized sensors. Three types of the industrial safety helmet were tested, one without ventilation openings, one with small ventilation holes made by the researcher, and one with manufacturing ventilation openings in the top region. Up to eight subjects were tested for each type of helmets in various ambient conditions. Results give the original measurements of the in-helmet micro climate. Results also show some interested interaction between the in-helmet temperature and relative humidity.

Keywords: comfort, industrial helmet, micro climate, perception, thermal/ moisture mapping.

1 Introduction

Construction or assembling of a manufacturing line is usually a physically demanding activity with workers working for periods of up to 8 hours in a day. In general, such activity could be even more demanding by the fact that workers are required to wear protective equipment in order to prevent serious injury that may result from being hit by falling nuts and/or bolts. Combined with the fact that construction or assembling are often carried out in the heat of summer, this may give rise to physiological and psychological strains for workers who have to wear helmets to protect their heads from injury by law. It is therefore vital to minimise any discomfort that may result from this equipment, in order to ensure the optimal conditions are available for the user. This discomfort may be either: thermal discomfort - in terms of heat and sweat, or mechanical – as a result of friction or irritation.

Research on strength of helmets under impact loading has been carried out extensively [1-4], since protection offered to users must be the most important safety issue for helmets. Various designs and material selections were made to maximise impact resistance of the helmet. An overall review of research on industrial helmet design was given by Proctor in early 1980s [5]. The review covered statistics on the incidence and severity of head injuries in accidents where helmets were being worn, as compared with those where helmets were not being worn. It also covered biomechanical evidence and helmet testing. However, the comfort dealt by the review was for contact mechanical pressure, although the review pointed out discomfort

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 678–686, 2007.

could also be caused by a high temperature inside the helmet. Therefore, thermal comfort was not a major issue at that time for industrial helmets, although research on how skin temperature affect on sweating and aerobic performance during severe work was carried out [6].

Thermal comfort has drawn attention and attracted some research since 1970s. Early research was undertaken by using modular liquid-cooled helmet liner in astronaut helmets to improve thermal comfort [7]. However, the serious research was only initiated early 1990s. Abeysekera and Shahnavaz [8] carried out ergonomics evaluation of modified industrial helmets to be used in tropical environments. Liu et al [9] undertook subjective evaluation in warm field conditions based on human perception and the corresponding sweating manikin head. Redesign of an industrial helmet was also carried out to improve its thermal properties [10]. In this research thermal couples were used to produce the temperature contour underneath the helmet in a manikin system. Furthermore, design recommendations with ventilation paths were made. More recently, Davis et al [11] undertook investigations how ventilation affect safety helmets in a hot environment. Psychophysical results showed that ventilation contributes to greater helmet comfort. In general, although there were some research activities on thermal comfort of industrial helmets, such research was not extensive, also only related to subjective physiological and psychophysical responses and/or manikin measurements. The only direct measurements of multi-point temperatures and the relative humidity (RH) were carried out inside a shoe for studying of the shoe climate [12 – 14].

In this paper, a test rig was developed to measure multi-point temperature and the RH inside an industrial helmet using micro sensors. Total thirteen locations were chosen to pick up measurements, which represent regions of the front, back, top, left and right of the helmet. There were three types of helmets tested, one without ventilation openings, one with small ventilation holes made by the researcher, and one with manufacturing ventilation openings in the top region. Up to eight subjects were tested for each type of helmets in various ambient conditions. Typical test duration was about 50 minutes, which was comprised of resting and walking periods on a treadmill. Experimental results show that the RH varies significantly inside a helmet whilst the corresponding temperatures change gradually depending upon ambient conditions. Typical variations of the temperature and the RH inside a helmet are 1~1.5 °*C* and 25~35% respectively for a relatively hot ambient temperature of about 30°*C*, whilst 2~3°*C* and 10~20 % respectively for a comfortable ambient temperature of about 24°*C*. These results show that even though the temperature inside a helmet does not vary greatly the corresponding RH can still change greatly. This gives scope to improve thermal comfort of a helmet since the RH is closely linked to it. Discussions on how to improve the thermal comfort are also given based on this pilot study.

2 Experimental Work

A test rig was developed to measure multi-point temperature and the relative humidity inside an industrial helmet using micro sensors. The test rig includes an industrial

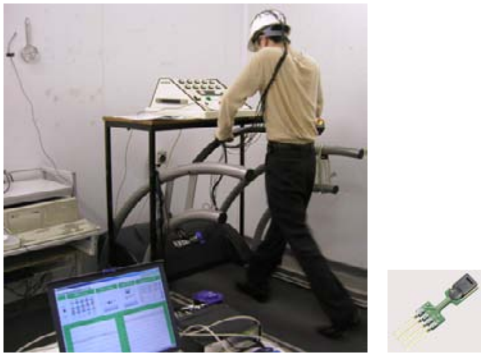


Fig. 1. The apparatus and subject during walking session

safety helmet with 13 sensors attached, a data logger, and a laptop. Two extra sensors were used to record the ambient conditions. Figure 1 shows an on-going test.

There were three types of commercially available industrial safety helmets tested. The Type I and Type II helmets were from the same white helmets except the Type II was provided with 22 small holes (3 mm in diameter) in the front, the side and the top areas. The Type III helmet was the red helmet that had 6 manufacturing ventilation openings sized 2 x 5mm each on the top region. Figure 2 shows the industrial safety helmets used in the test.

Type I helmet, Type II helmet, Type III helmet,

Fig. 2. Industrial safety helmets used in the study

The sensors used are capable of measuring the temperature and the relative humidity by using the same probe.

Total thirteen locations were chosen to pick up measurements, which represent regions of the front, the back, the top, the left and the right of a helmet. The selected positions were symmetrical, as shown in Figure 3. In order to embed the sensors, 13 holes with the diameter of 5 mm were drilled on the helmet. Sensors were placed above the inner surface of a helmet in a certain distance to avoid any direct contact to the head. In order to secure the position and maintain the original micro climatic conditions inside a helmet, all holes were sealed with silicone glue.

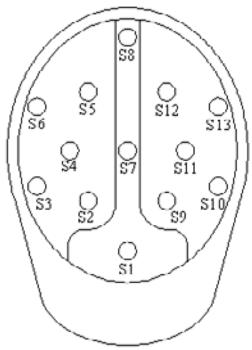


Fig. 3. The location of the sensors

The tests were carried out in the gymnasium at the Liverpool University Sport Centre. Up to eight healthy subjects were tested for each type of the helmets in various ambient conditions. All subjects were required to fill a consent form before carrying out the test. Typical test duration was about 50 minutes, which was comprised of resting and walking periods on a treadmill. Each resting and walking took about 10 minutes. The speed of tread mill was set at a walking speed of 5 km/h and measurements of the temperature and the RH were recorded every 5 seconds. 15 data were collected each times i.e. 13 from inside the helmet and other 2 from the ambient.

To asses the perception of the subject during the test, the thermal and moisture subjective perception forms and questionnaires were designed. The subjects were asked to fill both perception forms in every 10 minutes starting from t = 0 minutes. Both subjective perceptions are based on 5 scales. For temperature the scale are 1−cold, 2−cool, 3−neutral, 4−warm and 5−hot. The moisture scales are 1−wet, 2− slightly wet, 3−neutral, 4−slightly dry and 5−dry. The subjects were also asked to tick the area/areas where they felt hot, humid and uncomfortable for five areas, as shown in Table 1. All subjects were also asked to suggest the area/areas where should have openings.

Table 1. The breakdown of sensor areas

Area	**Sensor Locations**
Front	Sensor 1
Left	Sensor 2,3, 4,5 & 6
Right	Sensor 9,10, 11,12 &13
Top	Sensor 7
Back	Sensor 8
Ambient	Sensor 14 & 15

3 Result and Discussion

In general, when the helmet was put on, the temperature and humidity inside the helmet increased straight away. As the subject started to walk, both temperature and the relative humidity are decreased due to the dynamic air flow between the ambient

and the inside the helmet. Such air flow introduced efficient heat exchanges, which is good to the thermal comfort. The pattern continued until the end of the test.

Table 2 shows three typical ambient conditions and variations of the temperature and the relative humidity for the same position during the test for three types of the helmets. Clearly, for helmets with ventilation openings, more variations on both in-helmet temperatures and the RHs are indicated.

Table 2. The summary of the temperature and the relative humidity variations inside helmets during the tests

Helmet type	Temperature (ºC)			Relative Humidity (% RH)		
	Ambient	Variation (Static)	Variation (Walking)	Ambient	Variation (Static)	Variation (Walking)
I	25	0.5-1.3	0.5-0.7	50	2-23	2-10
II	22	0.5-1.4	0.5-1.8	55	5-20	5-24
III	29	0.5-1.5	0.5-1.2	45	5-23	5-28

In order assess the relative changes of the in-helmet temperature and the relative humidity against the ambient conditions after the measurement is stabilised, the corresponding parameters are defined as follows:

$$\Delta T(\%) = \frac{T_{\max} - T_{\min}}{T_{ambient}} \tag{1}$$

$$\Delta RH(\%) = \frac{RH_{\max} - RH_{\min}}{RH_{ambient}} \tag{2}$$

3.1 The Type I Helmet Without Ventilation Holes

The most of the subjects showed the relatively high and low temperatures at top/left and front/back areas, respectively. In terms of the relative humidity, most of the subjects showed the relatively high and low humidity at the top/front and the back areas, respectively. Figure 4 gives the temperature and the RH versus the testing duration for a subject, respectively.

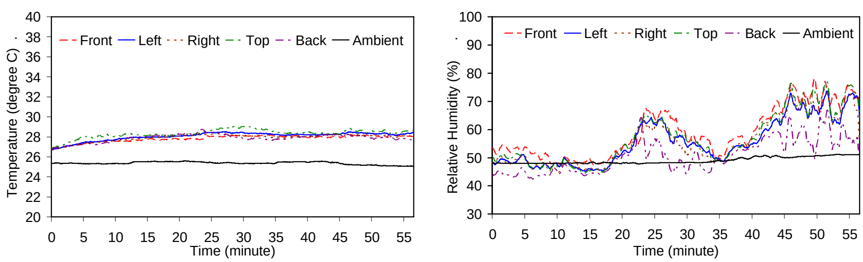
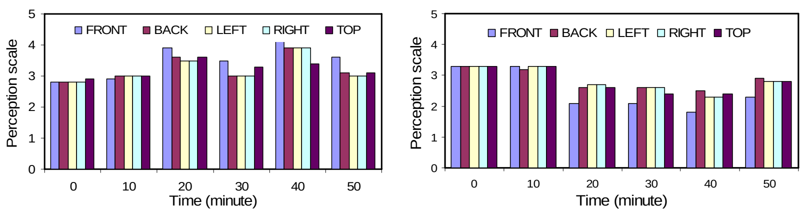


Fig. 4. In-helmet temperature and the RH versus test time for the Type I helmet

The type I helmet does not have any ventilation openings. There is hardly any heat/moisture exchange through the top region of the helmet, except the bottom circumferential air gap. Therefore, temperature measurements inside the helmet pretty close to each other. However, it is worth pointing out that the RH is building up through the testing period, also the change of the RH from location to location is significant. The relative changes of the in-helmet temperature and the RH against the ambient condition $\Delta T(\%)$ and $\Delta RH(\%)$ are 5.8% and 67.6%, respectively. The relatively high temperature measured on the top region is due to that the heat is moving up but there is no opening allowing it to escape. Also sweating on the top region is low which does not easy the temperature building up. On the contrary, the higher RH in the front caused by sweating may contribute the relatively low temperature measured, due to water vapour.

The subjective perceptions of the Helmet I show that the most hot and wet areas are front area, even though the temperature measurement does not support this. This may be blamed by the high RH. A humid local environment in the front could give the subject such feeling. Figure 5 shows the averaged thermal and moisture perceptions, respectively. These bar charts indicate that the hot thermal feeling coincides with the wet humidity feeling.



Thermal perception　　Relative humidity perception

Fig. 5. Averaged subject perceptions

3.2 The Type II Helmet with Small Ventilation Holes

In most of the cases, the left/right sides and the top areas show relatively high temperatures, whilst the front and the back relatively low temperature. However, the temperature in the back area tends to catch up, which gives the similar readings to that in the side in one or two occasions. The RHs demonstrate a general up trend for almost all cases. The RHs in the front and the back fluctuate the most, which are usually across over with each other through the walking and resting periods and also gives the upper and lower bound readings. Figure 6 shows typical measurements.

It can be seen that the RHs in the top and the back areas are decreased significantly during the walking period due to better air movement, also increased significantly during the resting period due to lack of air movement. In addition the chart indicates the coincidence between the temperature and the RH in some extent, i.e. the higher temperature corresponding to the higher RH. The relative changes of the in-helmet temperature and the RH against the ambient condition $\Delta T(\%)$ and $\Delta RH(\%)$ are 9% and 42.2%, respectively. The small ventilation holes seems having a certain effects on

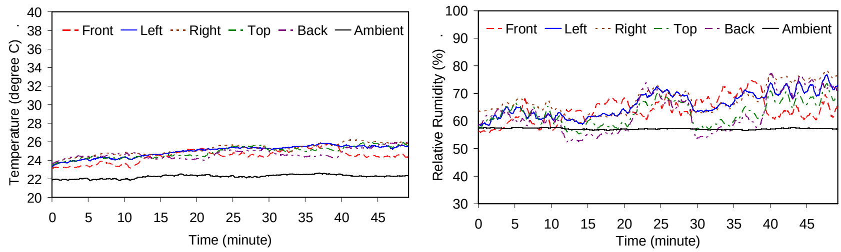


Fig. 6. In-Helmet Temperature and the RH Versus Test Time for the Type II Helmet

the in-helmet micro climate in terms of lower variation of the RH comparing with the Type I helmet. However, in the comfortable ambient conditions of 22°*C* and 57% RH, such effect is less significant.

Table 3 shows the summary of the questionnaire answer for the Type II helmet. The table shows that the majority of the subjects feel that the front area is the hottest, most humid, most uncomfortable and need openings. Such perception data can only be taken as a rough indication, especially for such small number of subjects.

Table 3. The summary of questionnaire answers for the Type II helmet

	Front	Back	Left	Right	Top	No comment
Hottest	5	1	0	0	1	1
Most Humid	6	1	0	0	0	1
Most Uncomfortable	3	1	0	0	3	1
Opening suggestion	3	3	0	0	0	2

3.3 The Type III Helmet with Manufacturing Ventilation Openings

The relatively high temperature regions are shown in the side and the back areas, with the relatively low temperature in the front and the top areas. For the relative humidity, most of the subjects give the relatively high RH measurements in the side and the back areas and the relatively low RH in the front and the top regions. Figure 7 shows typical measurements. The in-helmet temperatures are gradually increasing in the first

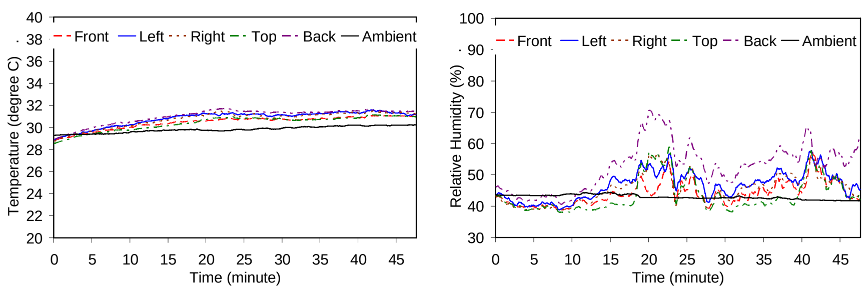


Fig. 7. In-helmet temperature and the RH versus test time for the type III helmet

22 minutes, and then reach an almost stable condition afterwards. The highest difference from location to location is only about 1°C. However, the pattern of the RH is quite different from that for the Helmet Type I. The high RH in the back area may be caused by no ventilation openings there, which helps the RH builds up. The general trend of the RH is almost along the horizontal axis. The highest relative changes of the in-helmet temperature and the RH against the ambient condition $\Delta T(\%)$ and $\Delta RH(\%)$ are 3.4% and 75% at about the 22nd minute, respectively. Such relative changes are gradually reduced. Big manufacturing openings on the top of the helmet do have some positive effects on the in-helmet micro climate. The highest RH recorded is about 70% for a high ambient temperature of 29 °C, which is less than that (77.5%) for the Helmet Type I.

Table 4 shows the summary of questionnaire answers. It is found that the majority of the subjects feel that front area is the hottest, most humid, most uncomfortable and needs openings. However, the perception is not quite coincided with the measurements. This may be partially caused by different ambient conditions.

Table 4. The summary of questionnaire answers for the Helmet Type III

	Front	Back	Left	Right	Top	No comment
Hottest	6	1	0	0	0	1
Most Humid	5	3	0	0	0	0
Most Uncomfortable	5	0	3	0	0	0
Opening suggestion	5	1	0	0	1	1

The results of this study show that although the in-helmet temperatures from location to location do not vary much, their patterns are still noticeably different for these three types of helmets. The in-helmet temperature varies in a relatively small amount for the Type I helmet due to no ventilation hole, whilst in a higher amount of change for the Type II and III helmets with ventilation openings. It is likely that ventilation openings contribute more air movement in the corresponding regions inside the helmet, which cause the temperature with a higher spread. More interesting phenomenon lays on the relative humidity. For the Type I helmet, as there is limited air exchange higher RHs are usually recorded in the front, the side and the top as expected. Also the changes of the RH in the walking and rest periods are great. With introducing limited ventilation openings to the Type II and III helmets, there are some in-helmet RH shifts. The RH in the back is clearly increased a lot and becomes the relatively high RH, and the RH on the top is decreased as certain amount of moisture can escape from the openings.

Although the test results show interesting phenomena, the measurements were carried out in uncontrolled ambient conditions. This gives difficult to compare results with each other. Future experimental work needs to be conducted in an environmental chamber with fully controlled climate conditions or in a less disturbed room with partial control of temperature and the relative humidity.

4 Conclusion

A test rig has been developed to obtain thermal/moisture mapping of industrial safety helmets. The rig can pick up both temperature and the relative humidity in 13 locations inside a helmet. Three types of the helmet, with and without ventilation openings, have been tested. Measurements reveal that the ventilation openings are crucial to regulate the micro climate inside the helmet, so as to the thermal comfort. The initial output may be useful to helmet manufacturers to consider where to make ventilation openings. More research needs to be undertaken on precise thermal/moisture mapping and link it to a more broad perception base, furthermore to computer modelling. However, these preliminary results pave the way to undertake further research leading to optimise thermal comfort of the helmet.

References

1. Proctor, T., Rowland, F.J.: Development of standards for industrial safety helmets. Occupational Accidents 8, 181–191 (1986)
2. Rowland, F.J., Patel, B., Keithley, J.H.: Some industrial and climbing helmets subjected to a range of impact energies. Occupational Accidents 10, 29–37 (1988)
3. Gilchrist, A., Mills, N.J.: Construction site workers helmets. Occupational Accidents 9, 199–211 (1987)
4. Mills, N.J., Gilchrist, A.: Industrial helmet performance in impacts. Safety Science 16(3-4), 221–238 (1993)
5. Proctor, T.: A review of research relating to industrial helmet design. Occupational Accidents 3, 259–272 (1982)
6. Davies, C.T.M.: Influence of skin temperature on sweating and aerobic performance during severe work. Journal of Applied Physiology 47(4), 770–777 (1979)
7. Williams, B.A., Shitzer, A.: Modular liquid-cooled helmet liner for thermal comfort. Aerospace Medicine 45(9), 1030–1036 (1974)
8. Abeysekera, J.D.A., Shahnavaz, H.: Ergonomics evaluation of modified industrial helmets for use in tropical environments. Ergonomics 31(9), 1317–1329 (1988)
9. Liu, X., Abeysekera, J., Shahnavaz, H.: Subjective evaluation of three helmets in cold laboratory and warm field environment. Industrial Ergonomics 23(1), 223–230 (1999)
10. Hsu, Y.L., Tai, C.Y., Chen, T.C.: Improving thermal properties of industrial safety helmets. Industrial Ergonomics 26(2), 109–117 (2000)
11. Davis, G.A., Edmisten, E.D., Thomas, R.E., Rummer, R.B., Pascoe, D.D.: Effects of ventilated safety helmets in a hot environment. Industrial Ergonomics 27(5), 321–329 (2001)
12. Covill D.G, Guan Z.W. , Pope D.J., Bailey M.: The effect of shoe fit on subjective responses, temperature and relative humidity levels. In: Proceedings of 4th International Conference on The Engineering of Sport, Kyoto, Japan, pp. 588–594 (September 3- 6, 2002)
13. Covill, D.G: Experimental and numerical modelling of shoe climate, PhD Thesis, University of Brighton (2005)
14. Raval, H., Guan, Z.W., Bailey, M., Covill, D.G.: Simplified 3-D FE model of thermal conditions inside a shoe. In: 9th International Conference on Advanced Computational Methods and Experimental Measurements in Heat and Mass Transfer, July 5-7, 2006. The New Forest, UK (2006)

Common Work Space or How to Support Cooperative Activities Between Human Operators and Machine: Application to Air Traffic Control

Benoît Guiost and Serge Debernard

Laboratoire d'Automatique et de Mécanique Industrielles et Humaines - UMR CNRS 8530
Université de Valenciennes et du Hainaut Cambrésis
Le Mont Houy, F-59313 Valenciennes Cedex 9, France
{Benoit.Guiost,Serge.Debernard}@univ-valenciennes.fr

Abstract: For several years, the need for air-traffic control has been continuously increasing. In order to maintain aircraft safety, different support tools have been built and tested. Our conviction is that it is necessary to conceive a more cooperative tool that would allow a "true team work" between air-traffic controllers and their support tools, by making the support tool part of the team rather than a substitute for air-traffic controllers. With a view to developing such a tool, we studied the cognitive activities of air-traffic controllers in a situation of cooperation involving two human operators. This paper presents a theory of human-human cooperation strategies based on our work, as well as our experimental protocol, analysis methods and observations.

Keywords: Air-traffic control, cognitive analysis, shared workspace, HCI, experimental protocol.

1 Introduction

Air-traffic control (ATC) is at the centre of a complex system in which the human operator holds the most important place. The principal task of air-traffic controllers is to ensure aircraft safety. In order to maintain security in the context of escalating air-traffic, one solution is to integrate assistance tools that regulate the controllers' workload. However, several studies performed at the LAMIH (Laboratoire d'Automatique et de Mécanique Industrielles et Humaines: Laboratory of Industrial and Human Automation, Mechanics and Computer Science) have revealed decision-making conflicts between the support tool and the human operators [8], because the tool and the controllers have different decision-making strategies. To solve this problem, we have begun to work on a more cooperative kind of support tool system. This paper presents the models which were used to specify a support system for the air-traffic controllers. Then it presents the experiments and the analyses, which were performed to assess the human-machine system.

2 Human-Machine Cooperation

Our study is based both on a theory of human-human cooperation (HHC) and on a theory of human-machine cooperation (HMC). Section 2.1 defines HHC and HMC.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 687–696, 2007.

Section 2.2 presents a model of human operator, and section 2.3 explains how those concepts were used to create a theory of HHC.

2.1 Definition of the Human-Machine Cooperation

Hoc defines the cooperation between two agents in this way [7]:

" Two agents are in a cooperative situation if they meet two minimal conditions:

- *Each one strives towards goals and can interfere with the other~~s~~ on goals, resources, procedures, etc.*
- *Each one tries to manage the interference to facilitate the individual activities and/or the common task when it exists.*

The symmetric nature of this definition can be only partly satisfied"

The goals described above are not the ones set by the supervision and control process but rather those for accomplishing a particular task. Interferences are interactions between the activities of several agents. Their nature can be positive or negative. "Positive interaction" refers to normal interaction between agents, and "negative interaction" refers to conflict between agents. The objectives of the HMC are to assist human operator in order to increase the system's performances while avoiding a human overload. So it is necessary to minimize negative interactions between operators and support tools. The HMC can be defined from structural or functional aspects:

Structural Human-Machine Cooperation can take two forms: vertical or horizontal [9][10]. In a vertical structure, support tools can't ever act on the system. Only operators have this capacity. However, because the support tools have the same information on the process as the human operators, they can give to the operators some advices. On the other hand, in a horizontal structure, both the human operators and the support tools can act on the system. The support tools have reasoning capacity in real time, which puts the two decision-makers (human and machine) at the same hierarchic level. Different tasks are allowed allocated between agents in an attempt to regulate the workload of human operators.

Functional Human-Machine Cooperation can take three forms: augmentative, debative and integrative [15]. In the augmentative form, all agents have the same abilities and work together to complete one task that is too extensive for a single operator alone. This task is divided into several sub-tasks that are distributed among agents. In the debative form, agents again have the same abilities, but rather than share work on one task, they must compare their individual results on order to obtain the best solution. In the integrative form, agent abilities vary, and the task is divided into sub-tasks that are allocated according to these abilities. The different agents supplement one another as they seek to accomplish the task. Each operator contributing his/her part towards task completion. The three forms of cooperation give indications on various mechanisms of co-operation when two human operators work together.

The above forms and structures are used in the conception of a support system for human operators. The following section presents the internal structure of HHC. This structure, once integrated into Hoc's definition (cited above), will provide the basis for the definition of our future system.

2.2 The Rasmussen' Model for Human Activities of Problem Resolution

Rasmussen [13] has presented a model that describes the human cognitive activities for problem resolution. Firstly, the human operator collects all the data allowing him/her to identify an abnormal situation. Then, those information allow the human operator to perform a diagnosis of the situation identifying the problem precisely. When the problem is identified, the operator can build a schematic solution answering the problem according to the inherent constraints with the system. This stage leads to the development of the solution in terms of goals and under-goals, and finally to the implementation of the solution.

The model of Rasmussen [14] revealed three classes of human operator behaviours:

- The human operator applies the skill-based behaviour when it is confronted to a known situation and whose solutions are applied automatically and spontaneously. This behaviour is related to automatisms.
- With the rules-based behaviour, the human operator detects a known abnormal situation and proposes a solution resulting from rules or preset procedures that she/he memorized beforehand.
- Lastly, the knowledge-based behaviour is applied when the human operator is confronted with an unknown situation and that she/he uses his knowledge and its experiment to invent an adequate solution.

2.3 The Common Frame of Reference (COFOR)

To accomplish a task, human operators build themselves a frame of reference within which they represent the process and the process state. Thanks to this construction, human operators can plan their actions and detect abnormal evolutions of the process [11]. Within the context of HHC (cooperation between human operators), agents exchange information verbally about their individual frames of reference in order to build a common frame of reference (COFOR). They use this COFOR to accomplish tasks whose goals or sub-goals are linked. The COFOR becomes a reference between operators and thus corresponds to Pacaux-Lemoine and Debernard's definition.

In order to minimize and optimise the number, duration and contents of these exchanges, agents often construct their own representations of what they believe a colleague's COFOR to be; however, sometimes these interpretations are partially wrong. To improve cooperation between human operators and the future support tool, a more formal method for constructing a COFOR would seem to be necessary [2]. One suggested method is called Common Work Space (CWS).

Debernard and Hoc have proposed defining the structure and function of support tool in terms of CWS notions and Schmidt's three forms of cooperation. The three forms are implemented in this way [1]:

- Debative form: agents add the data they judge significant to the CWS. When interferences appear, they must negotiate to eliminate them.
- Integrative form: one agent adds data to the CWS. Other agents access the information necessary to accomplish their tasks.
- Augmentative form: the agents themselves decide Task sharing and data addition methods.

3 A Model of Human-Machine Cooperation Through a Common Workspace

Certain processes, as the air-traffic control must be automated partially. When the automation goal is to decrease the human operator workload through an assistance to perform some task, it is necessary to define the function allocation between the different agents. If the function allowed to human operators are related to those allowed to the support system, it will be necessary to provide the various agents, the means to cooperate.

With a synthesis on the structures and forms of the Man-machine cooperation, on the common frame of reference and the model of Rasmussen for problem resolution, a co-operative Human-Machine system can be modelled (Figure 1).

The COFOR of human operators is placed between the artificial and human agents. Consequently, they can build a common representation of the situation. The hardware representation of the COFOR is called Common Workspace. This Common Workspace is composed of the data generated by the activities of problem resolution defined by Rasmussen. In addition, this model allows to represent the function allocation between the agents and to highlight the various forms of cooperation within the Schmidt forms. For example, the support system and the human operator can add regularly new information to the common workspace (Figure 1 - Step 1). It is an augmentative co-operation. The human operator can define the problem. This definition is integrated by the support system that deduces the strategy that is able to solve the problem (Figure 1 – Step 2). It is an integrative co-operation. The human operator and the support system can provide the solution in order to compare their results (Figure 1 - Step 3). It is a debative co-operation.

This model was used to specify a co-operative support system for the air-traffic controllers. The specification of this support system led to the development of an experimental platform called AMANDA.

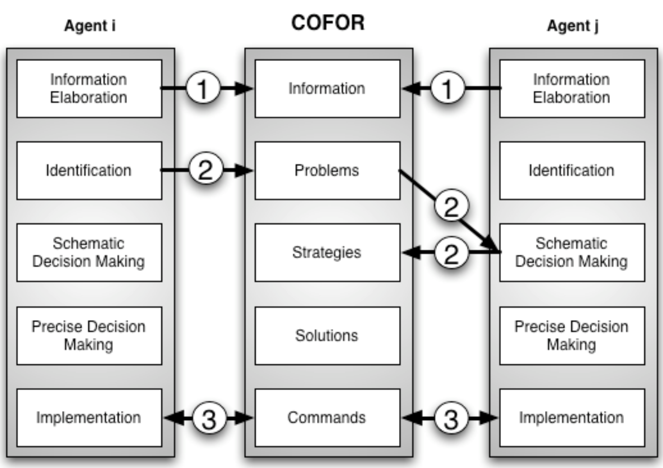


Fig. 1. Cooperation through a Common Frame of Reference

4 Design of a Man-Machine System for the Control of the Air Traffic

French ATC is a public service in charge of the safety and regulation of air-traffic. There are two types of control: approach control and “en route” control. Approach control concerns the preparation and supervision of landing aircraft. “En route” control manages all the other traffic, overseeing any aircraft flying over French air space. Our study deals with “en route” control. French air space is divided into five control centres, each containing several sectors according to air-traffic density. Air space is also divided into several flight levels in order to increase capacity. When the trajectories of two or more aircrafts cut across the same point on the same flight level, a situation for conflict exists. To resolve the potential conflict, controllers must separate aircrafts by sending them different orders (heading, flight level, speed).

Working controllers perform their tasks in groups of two: a planning controller (PC) and a radar controller (RC). PCs have a strategic role; they detect eventual conflicts and inform the RC as well as coordinating their own sector activity with that of all adjacent sectors. RCs have a tactical role; they supervise the air-traffic and find solutions for eventual conflicts. Each controller has radar image, which displays in real time information about the aircrafts in the controlled sector and paper strip that indicate some information as speed, destination, and flight level… Our experimental platform attempts to place controllers into the most realistic situations possible in terms of the different elements of their real tasks and their control post.

The goal of the AMANDA project is to propose to the air-traffic controllers a support system able to integrate their situation representation and to aid them in their search for solution to ensure the aircraft separation. This project was carried out according to three steps. The first step corresponds to the identification of the contents of air-traffic controllers COFOR through a study of their cognitive activities. The second step uses the result of the first in order to design the Human-Machine system. The third step corresponds to the evaluation of the Human-Machine system

4.1 Identification of the Contents of the Common Reference Frame

In order to identify the contents of the RC's COFOR, the experimental platform included one PC and two RCs assigned to control the same traffic. Aircraft were divided between the two RCs, neither of which could act on an aircraft that was not assigned to him/her. All the potentially conflicting aircraft were not assigned to the same controller, insuring that the two RCs had to cooperate implicitly in order to solve conflicts.

In France, air-traffic controllers handle, on average, twenty-five aircraft per hour. In order to insure that the controllers would cooperate, the scenarios used in the protocol imposed a control of fifty aircrafts per hour, thus putting controllers into a realistic, though overloaded, situation.

This study shows that more half of the air-traffic controllers’ activities (58%) relate to the COFOR construction and maintaining [4]. To build their frame of reference, the controllers identify the air conflicts as problems. These problems contain the conflicts planes, their altitude, the strategies of problem resolution and the instructions sent to the pilots. The strategies are often expressed by heading instructions and in forms explicit or implicit. An explicit strategy is composed of the two planes that must be

separated while an implicit strategy is only composed of the one plane that must be deviated of its trajectory to solve the conflict. The definition of the problems concerns an activity of imposition on behalf of the controllers, i.e. they add spontaneously information on the COFOR with mutually confidence. On the contrary, the controllers have a strong activity of negotiation on the strategies. The strategy of conflict resolution is a crucial point in the process of problem resolution. Then, the technical solutions and their application rise naturally from the selected strategy and are not any more the object of negotiation by the controllers.

The results of this study made it possible to specify the Human-machine system, the function allocation and the modes of co-operation between the agents.

4.2 Specification of the Man-Machine System

Figure 2 shows the function allocation between the controllers and the support system as well as the modes of co-operation. The COFOR is materialized to a graphic interface placed between the agents. The support system, called STAR, is not able to perform all the activities of problem resolution.

The human operators and the support system can add information to the common workspace. The co-operation is an augmentative form. The air controller must carry out the definition of the problem, and controls the overall representation of the traffic. However STAR will be able to complete the beforehand the problem representation. These activities concern activity of identification of the Rasmussen model. The co-operation is a debative form.

In order not to cause decisional conflicts between human and machine, STAR will not be able to make strategic decision. Consequently, STAR must be able to integrate the strategy of controllers thanks to the common workspace and to calculate the regarding solution. The co-operation is an integrative form.

Human operators or STAR can carry out the command. The human operator carries out the command allocation between the agents. This command must enable her/him to control its workload. This command is called delegation because it relates to a very precise micro-task and not on the whole management of conflict.

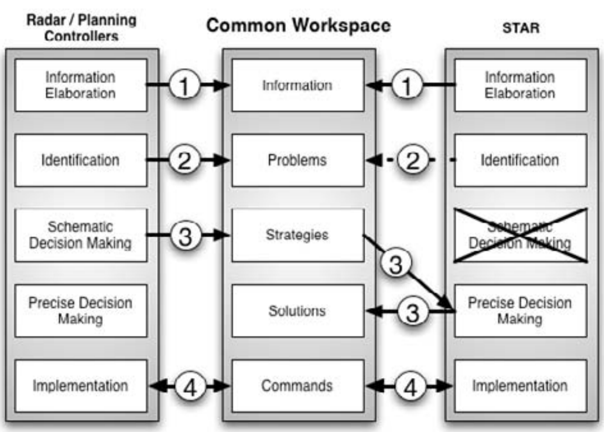


Fig. 2. Function Allocation and cooperation between Planning and Radar Controller and STAR

4.3 Interfaces of the Man-Machine System

The common workspace has been implemented that allows human controllers to cooperate both with the system and with their colleagues.

This common workspace is composed of a graphic interface, on which controllers can display conflicts as “clusters” composed of the aircraft in conflict, the minimal distance between these aircraft, and the order sent to the aircraft either by the controllers or the support system. Controllers can also define a strategy for solving a conflict and ask the support system to produce a solution, which the controller can then accept or decline. The management of the contents of each cluster is carried out on a specific interface that covers totality of the common workspace when one double-clicks on the cluster to manage. This man-machine system was confronted with professional air controllers during a third phase of evaluation.

4.4 Evaluation of the Man-Machine System

In order to assess the impact of this type of support system on controller activity, six professional controllers tested the AMANDA V2 experimental protocol. Each pair of controllers worked with the new support system for two days, in two phases. In the first phase, they familiarized themselves with the system, participating in several training scenarios that allowed them to gradually discover the various system functions. In the second phase, the controllers used the support system in three different operating modes, selected with the aim of evaluating the effectiveness and the relevance of the various modules of the tool.

In the first mode (situation A), the STAR computational solution tool was disconnected, and the two human operators had to rely on the common workspace for any cooperative activities. However, in the second and the third mode, the entire support system was available, and the goal was to determine the best way to share the various tasks between the planning controller (PC) and the radar controller (RC). In the second mode (situation B), the PC was responsible for building the clusters and defining strategy, and the RC had to decide whether to accept, modify or refuse the strategies defined by his/her colleague. If the RC accepted the strategy, he/she could then choose to delegate it to STAR or to send the instructions directly to the pilot. On the other hand, if the RC refused his/her colleague's strategy, then he/she could define a new strategy and delegate its application once the STAR calculation showed the new strategy was appropriate. In the third mode (situation C), the PC was still responsible for building the clusters, but it was the RC who defined the problem-solving strategies and delegated them, if necessary.

The analysis was based on different criteria extrinsic and intrinsic [5]. The workload evolution was studied according to the situations and was measured with the TLX method [6].

According to the TLX questionnaires, the controllers find that the best compromise between support tool availability and workload was provided by situation B. This situation was designed to test a task distribution schema that shares tasks between the planning controller and the radar controller. The situation-B conditions required the planning controller to fully fulfil his/her strategic role by defining both the clusters and the problem-solving strategies. This situation did not provoke a significant rise in

the controller workload, compared to situation A in which the controllers operated without assistance.

The study of the objective criteria, i.e. the variables measured at output of the human-machine system, shows that the common workspace allowed the controllers to build their COFOR and to share a representation of the situation with the support system STAR. Moreover, the temporal constraints induced by STAR to obtain a good solution have obliged the air controllers to increase their anticipation of the problems.

Concerning the use of STAR, controllers have delegated explicit strategies as "TO PASS BEHIND". The utilisation ratios of STAR according to the situations have significant differences. Those different can be related to the air controllers workload measured with TLX. Indeed, when it is the planning controllers who define the strategies (situation B), in 85% of the cases, she/he has affected a directive to solve the problem. This rate is only 63% when it is the radar controller who must define the strategy (situation C). When these problems were delegated, STAR was to apply two instructions to separate the flights: a heading instruction to deviate the flight of its trajectory then an instruction into “direct” in order to replace the aircraft on its trajectory. It is in situation C that STAR generally could begin separation (56% of the problems against 41% for the situation B). However, it is in situation B that the controllers have to generally let STAR apply the handing-over into direct (34% against 19% in situation C). The assumption is that the radar controllers are overloaded in situation C and that they wanted more quickly to give the flight into direct in order to be able to discharge from the problems. The objective measures confirm that it is the planning controller who must define the strategies of problem resolution.

5 Conclusion and Perspectives

The AMANDA project was realized with three phases. The first phase consisted of an exploratory step making it possible to carry out a model of the air-traffic controller common frame of reference and to identify their cognitive activities. The result results of the first phase were used to specify the function allocation between the human operators and the support system. Then, the function allocation was used to specify the modes of co-operation, which this is necessary to support a man-machine co-operation as effective as possible. These specifications were used to design a common workspace whose contents are managed by the human operators and STAR. This common workspace made it possible to the controllers to share a common representation of the situation with STAR and to maintain their situation awareness as defined by Endsley [3]. I.e., the interface presents the situation from the air conflict point of view but also allowed them to easily visualize problem allocation between human operator and STAR.

This study shows the impact of new support system on the function allocation between the planning controller and the radar controller who has been modified in order to as well as possible control the controller’s workload. Thus, the use of a tool able to integrate the strategy of the controllers to calculate the corresponding solution implies, on behalf of the controllers, anticipation more important than they have the practice of it. This anticipation must be managed by planning controllers, who work

upstream of the radar controller. This strategic decision-making carried out by the planning controller must make it possible not to overload the radar controller.

The AMANDA project begins a new phase, which consists in extending the common workspace to obtain a multi-sector co-operation an air-ground co-operation. The co-operation multi-sector must make it possible to broadcast the strategies between two sectors and thus to decrease uncertainties on the flights trajectories before it enters in a sector. The air-ground co-operation must make it possible to delegate on board the solution calculation. The assumption is that the embarked instruments will be more powerful to solve the problems than the grounds equipment, which work with uncertainties on the trajectories more important than on board.

References

1. Debernard, S., Hoc, J-M.: Designing Dynamic Human-Machine Task Allocation in Air Traffic Control: Lessons Drawn From a Multidisciplinary Collaboration. In: Smith, M.J., Salvendy, G., Harris, D., Koubek, R. (eds.) Usability evaluation and Interface design: Cognitive Engineering, Intelligent Agents and Virtual Reality, vol. 1, pp. 1440–1444. Lawrence Erlbaum Associate Publishers, London (2001)
2. Debernard, S., Cathelain, S., Crévits, I., Poulain, T.: AMANDA Project: Delegation of tasks in the air-traffic control domain. In: Blay-Fornarino, M., Pinna-Dery, A-M., Schmidt, K., Zaraté, P. (eds.) Cooperative Systems design (COOP'2002), pp. 173–190. IOS Press, Amsterdam (2002)
3. Endsley, M.R.: Automation and situation awareness. In: Parasraman, R., Mouloua, M. (eds.) Automation and human performance: Theory and application, pp. 163–181. Lawrence Erlbaum Associates, Mahwah (1996)
4. Guiost, B., Debernard, S., Millot, P.: Definition of a Common Work Space. In: 10th International Conference of Human-Computer Interaction, vol. 3, pp. 442–446,Crete, Greece (July 2003)
5. Guiost, B., Debernard, S., Poulain, T., Millot, P.: Task Allocatio In Air Traffic Control Involving A Common Workspace And A Cooperative Support System. In: IFAC ASBHS 2006, Nancy, France, Mai (2006)
6. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human mental workload, pp. 139–183. Elsevier, Amsterdam (1988)
7. Hoc, J.M.: Supervision et contrôle de processus - La cognition en situation dynamique. University' press of Grenoble, France (1996)
8. Lemoine, M-P., Debernard, S., Crevits, I., Millot, P.: Cooperation between humans and machines: first results of an experimentation of a multi-level cooperative organization in air traffic control. Cooperative Supported Cooperative Work. The journal of collaborative computing 5(2-3), 229–321 (1996)
9. Millot, P.: Supervision des procédés automatisés et ergonomie, Traité des nouvelles technologies Série automatique, Editions HERMES. Paris (Décembre 1988)
10. Millot, P.: La supervision et la coopération Homme-Machine dans les grands systèmes industriels ou de transport, Colloque national GIS Sciences de la cognition: Sécurité et cognition, Paris (Septembre 1997)
11. Pacaux Lemoine, M.P.: Coopération Hommes-Machines dans les procédés complexes: modèles techniques et cognitifs pour le contrôle de trafic aérien. Thesis. University of Valenciennes and Hainaut-Cambrésis, France 1996 (1998)

12. Pacaux-Lemoine, M.P., Debernard, S.: Common work space for human-machine cooperation in air traffic control. Control Engineering Practice 10(5), 571–576 (2002)
13. Rasmussen, J.: Notes on human error analysis and prediction. In: Apostolakis, G., Garriba, S., Volta, G. (eds.) proceedings of the NATO advanced study institute on synthesis and analysis methods for safety and reliability studies, pp. 357–389. Plenum Press, New York (1980)
14. Rasmussen, J.: Skills, rules, and knowledge; signals, signs and symbols, and other distinctions in human performance models. IEEE Transactions on systems, man cybernetics SMC13(3) (May/June 1983)
15. Schmidt, K.: Cooperative Work: a conceptual framework. In: Rasmussen, J., Brehmer, B., Leplat, J. (eds.) Distributed decision-making: cognitive models for cooperative work, pp. 75–110. John Willey and Sons, Chichester (1991)

Human Performance Enhancements: From Certification to HCI Innovation

Peter G.A.M. Jorna

National Aerospace Laboratory (NLR)
Anthony Fokkerweg2, 1059CM, Amsterdam
jorna@nlr.nl

Abstract. This paper addresses the need for human performance enhancements in relation to growth of air travel and keeping accident levels low. Performance can be enhanced by certifying equipment with Human factors based methodologies in order to assure compatibility of systems with human intricacies. Objective eye scan data can be used to facilitate compliance with new Flight deck certification rules. The intelligence of machines or computers cab be improved by allowing machine access to eye derived data as illustrated and discussed. With this technology it is possible for computers to become a true COMPANION for its user.

Keywords: human performance, certification, human machine interaction, HCI innovation, adaptive systems.

1 Safe and Reliable Human Performance

The societal demand for air travel has been consistently rising and this trend is expected to continue even with increasing fuel prices etc. Given the present levels of safety, there is a general concern for an increase in the absolute number of accidents. In order to keep air travel safe, there is a need to enhance safety in order to prevent this from happening. The human operators in the aviation system are often mentioned by accident investigators as a significant factor contributing to accidents. There even used to be a trend to stop the investigation after identifying human error as a cause of the accident. Fortunately, it has been increasingly recognised that human performance is not simply a matter of legal responsibility (in our company it is not allowed to make errors) but more a matter of humans interacting dynamically within a task and working environment, where, given the information available on the task at hand, behaviour made perfect sense from their perspective at the time.

Human error is not a cause for a system breakdown, but one of its symptoms

Flight crew are very motivated to survive their flights and will generally not act outside of their " human performance envelope", unless forced or provoked to do so. Examples of being forced into trouble, are driving too long without adequate rest opportunities and/or time pressure imposed by economic reasons, working in an ill designed distracting and fatiguing environment, having to use confusing documentation,

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 697–704, 2007.

work with inadequate tooling etc. The design of tools, documentation and software is known to sometimes provoke behavior that is not in line with the overall goal of achieving a task or mission. An example is easily found in day to day life. I just installed a new car radio and these devices have so-called pre-selects for easily finding your favorite stations.

And…… to make life even more pleasurable, they added a single button that initiates a complete reprogramming of all pre-sets. Unfortunately, they located that button as number one in a row of pre-selects (yes, pre-select one is actually in position two…..). The result is that I always have to remind myself of this design feature and if not, I will be punished by a complete reprogramming. Despite excellent motivation, years of professional training and a good mood in general, I do fail sometimes.

Design therefore can be a factor that influences the outcome of perfect normal behavior otherwise. I expect the first button of the pre-sets to be at the first position. Cleary this design is incompatible with my expectations based on a long life of learning and surviving (yes, there is good equipment around). But individuals have different lives and as a consequence also different expectations, such as cultural stereotypes influencing the use of faucets. With new systems it therefore makes sense to "validate" the expected or designed use of equipment by exposing it to various users. Or even better, put that requirement into official legislation.

2 Human Factors Certification as a Step Forward

The advent of aircraft with automation support on many levels, contributed to improved safety but it also initiated a new class of human behaviors that proved incompatible with the design, i.e. automation induced. Errors, as in case of my car radio. The FAA and JAA initiated a joint team to tackle these problems for future designs and aircraft updates. I was fortunate to participate and contribute to this regulation of Human factors Integration in flight deck systems. The challenge for this Human Factors Harmonization Working Group (HF-HWG) was formulated as such:

'Flight crews make a positive contribution to the safety of the air transportation system because of their ability to assess continuously changing conditions and situations, analyze potential actions, and make reasoned decisions. However, even well trained, qualified, healthy, alert flight crewmembers commit errors. Some of these errors may be influenced by the design of the systems and their associated flight crew interfaces even with systems and associated interfaces that are carefully designed. Although most of these errors have no significant safety consequences, or are detected and/or mitigated in the normal course of events, accident analyses show that flight crew performance and error have been identified as significant factors in a majority of accidents involving transport category airplanes. Accidents most often occur as a result of a sequence or combination of errors and safety related events (e.g., equipment failure, weather conditions). These analyses also show that the design of the aircraft flight deck and other systems can influence the performance of flight crew tasks and the occurrence and effects of some flight crew errors.'

After an intensive process, the team succeeded in formulating a legally acceptable text for rulemaking in this area. That text consolidated a umber of very important steps forward in the thinking about the "why" of human errors. For information look at http://www.researchintegrations.com/hf-hwg/

2.1 FAR/CS§ 25.1302 Installed Systems and Equipment for Use by the Flight Crew

This section applies to installed equipment intended for the flight crewmembers' use in the operation of the airplane from their normally seated position on the flight deck. This installed equipment must be shown, individually and in combination with other such equipment, to be designed such that qualified flight crewmembers trained in its use can safely perform their <u>tasks</u> associated with <u>the intended function</u> by meeting the following requirements:

(a) Flight deck controls must be installed and information necessary to accomplish these tasks must be provided.
(b) The flight deck controls and information intended for the flight crew use must:

i. Be presented in a clear and unambiguous form, at resolution and precision appropriate to the task, and
ii. Be accessible and usable by the flight crew in a manner consistent with the <u>urgency, frequency, and duration</u> of their tasks, and
iii. Enable <u>flight crew awareness</u>, if awareness is required for safe operation, of the effects on the aircraft or systems resulting from flight crew actions.

(c) Operationally-relevant <u>behavior of the installed equipment</u> must be:

i. Predictable and unambiguous, and
ii. Designed to enable the flight crew to intervene in a manner appropriate to the
task.

(d) To the extent practicable, the installed equipment must enable the flight crew to <u>manage errors</u> resulting from flight crew interaction with the equipment that can be reasonably expected in service, assuming flight crews acting in good faith. This subparagraph does not apply to skill-related errors associated with manual control of the airplane.

Several interesting elements can be found in the text of this rule (as marked by underlining.[1]). First, there is a clear and unique reference to the actual tasking of the crews and their working context, including other tasks and equipment. Secondly, it addresses the intended function of equipment, meaning that certain equipment is not to be used differently as intended for by its design. This is of legal importance as

[1] Note; the 25.1302 rule discussed here is not yet officially published and this could not be the definitive one.

manufacturers could else be held liable for all kinds of creative or ‘unexpected behaviors’ and that would essentially force them to block all innovation and that would be quit counterproductive. Note that the text does not describe (or prescribe) how to design a particular system like an FMS (Flight Management System) as that could also inhibit innovation (other ways of accomplishing that task or system function). The crew should also be enabled to build awareness by providing information and feedback on actions taken and not to be taken. Feedback is therefore required on ‘system behavior’. The use of this term in a legal document is actually a recognition for human factors as technical systems normally operate by ‘functional logic’. But people interpret that logic as behavior (some machines often behave erratic…) and react to it as they would react to living individuals (when was the last time you yelled to a computer controlled device..?). Try a new car radio!

The importance of this human factors rule is twofold. First, it will now be mandatory to validate the human behavior with new systems before production or the accident. Secondly, it’s a major milestone in acceptance of and need for integration of Human Factors in system design and operation.

3 An ‘Eye’ for New Behavior with(in) New Systems

A focus on the Human Factor has clear implications for any approach for setting up an evaluation method. Technical evaluations naturally seem to focus on specific equipment or ‘boxes’ on the aircraft that serve a function, while humans are on board to perform certain ‘tasks’ operating more than one piece of equipment. It is now mandatory to address the flight deck as an integrated working environment. Most tasks for flight crew are heavily dependent on visual information presented on displays. The investigation of *information scanning based on eye-data* is a good tool to use in evaluations. The use of eye based data is attractive as it reveals the strategies that are being used by the flight crew to access the visual data available on new systems in a flight deck. The most basic technique shows the eye scan superimposed on an image taken by a head mounted camera that films the ‘scene’. This technique is nice for on the spot applications, such as debrief during training, but it is to cumbersome and expensive for quality evaluations. A better technique is to combine the registration of eye movements, with tracking of head position as well. This will allow the use of point of gaze measurements that will directly show the position of the eye scan referenced to the actual equipment (For a review see Jorna & Hoogeboom 2006).

Our lab has extensive experience with these techniques and its known to reveal unexpected findings. One example is the potential response of human operators to software tools that were designed to assist. When task load increases, software tools should help to maintain acceptable individual workload levels. This is not always the case as illustrated below.

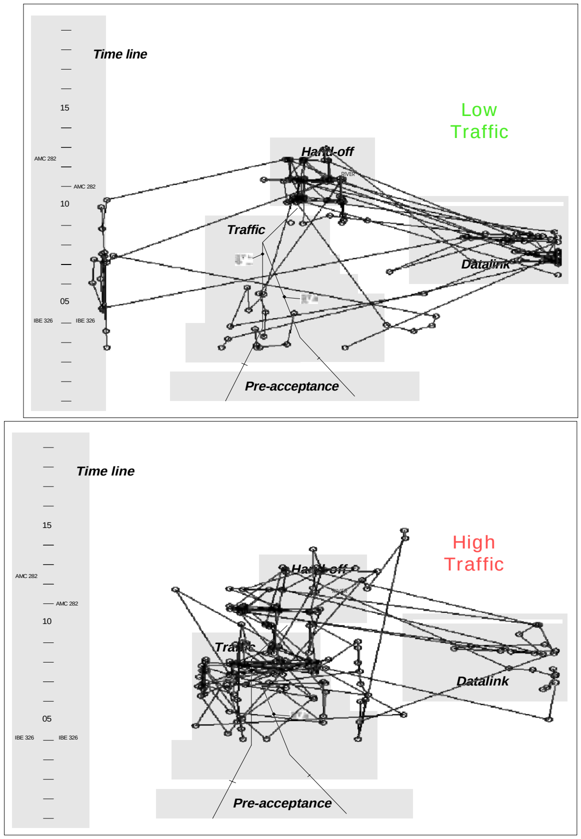


Fig. 1. Eye scan of an Air traffic Controller supposed to use the software tool Arrival manager as depicted by the bar on the left side, but drops the tool all together when the traffic gets tough

The Eye tracking technology itself has moved forward from the use of cumbersome, but very accurate, head mounted systems with camera's, mirrors and infrared sources, towards stereo camera's using video imaging and facial mapping. Easier to use, but less accurate so far. An illustration of a simulator cockpit application is provided. The point of gaze measures can be used for off line analysis of pilot scanning and information processing behaviour, but can also be generated in a near real time fashion. This opens an opportunity for innovating HCI by telling the machine what the user is seeing (or not…..).

4 Innovating Machine Intelligence

Based on the power of eye tracking data for revealing the 'state of the user', we invented the COMPANION HCI concept (Cooperative Observing MMI for Personalized Assistance and Narration as Induced by Operator Needs) .in which we allow the machine or computer to access the real time data obtained from the user by an eye

tracker system (Jorna, IOP MMI99002A project Ministry of Economic affairs). This opens the way to solve a major problem in HCI, i.e. the problem of computers not being very good in 'social intelligence'. Some examples: if you want to talk to your boss and he/she is on the phone, its natural to 'not start talking' until the phone call has finished. Computers don't, they simply interrupt. A companion equipped system will know that you are busy reading something on the screen and will as a consequence, hold the new Email (depending on its significance) until your eye scan reveals that you are free(er) to receive it. Another example is that you want to deliver an important message and you see that your friend is not there. You will try to bring the message to him. Computers don't. In a companion equipped aircraft the message will simply be moved to the display (head up, head down, visor, flight bag etc.) that you are looking at. The system knows that by its access to eye data.

Computers are no team players and bare any social intelligence, machine access to eye data can solve that

When the machine has access to eye data it can assess so-called 'states" of the user. In simple terms states such as: present/not present, busy or free, loaded/relaxed, reading or searching etc. Based on these (generic!!) states the computer can modify its behavior by changing the timing of displayed information, its location or even its format by shortening its content for a quick view. Work is now in progress within the HILAS project to develop a flight deck based application that can serve as a safeguard against missing any information by the crew. The avionics know what they are displaying, with a companion system included, they can assess if the crew has actually attended it….The military version is developed under the name ACS (Adaptive Cockpit Systems).

Fig. 2. Simulator tests for HCI research into *Pilot Observing Avionics* , using the Companion principle

4.1 From 'Looking' Towards 'Seeing'

Knowledge about eye focus positions on hardware or computer displays can be very informative and revealing, but it is not a guarantee that the information is actually being processed. You could be thinking about buying a new car radio, for instance…….This issue is addressed by adding information about level and changes in activity with respect to pupil size and/or heart rate. For practical purposes, pupil size has the advantage of being available through the same equipment measuring eye scanning. When information is actually being processed, the size of the pupil will increase (given that light levels do not change dramatically). To illustrate this process we show you the data on an experiment involving processing of auditory information only and auditory information combined with visual tracking (Hoogeboom & Jorna 2005). Pupil size is sensitive to mental load and can reveal valuable information in a variety of conditions as indicated.

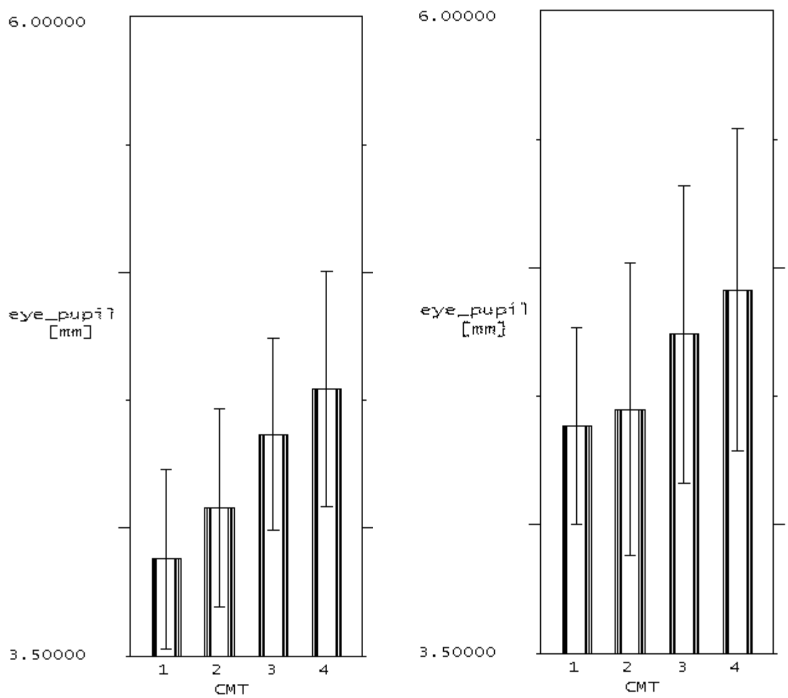


Fig. 3. Pupil size increase as a function of memory size (1-4 target letters) in an auditory Continuous Memory Task (CMT) and combined with visual tracking as depicted on the right side

5 Concluding Remarks

The integration of Human factors into aviation systems has taken some major steps forward. Its inclusion as part of the airworthiness regulations clearly confirms its importance and acceptance. Complying with these regulations will be more complicated

but new techniques are available that can help to harness the certification process into an affordable and not too time consuming process. The use of eye data has proven to be valuable in analyzing behavior with new (and old) systems. Its value can be increased even further by allowing the machine to access the eye data from its user. With this information the computer can evolve to a true Companion that will safeguard against overlooking critical information or overloading the user with data when you are not in a state to process at this instance. Industries who are interested in developing applications of this concept are invited to contact us for details.

References

1. Jorna, P.G.A.M., Hoogeboom, P.J.: Evaluating the Flight deck. In: Harris, D. (ed.) Human factors book on flight deck design, Ashgate, UK (2006)
2. Hoogeboom, P.J., Jorna, P.G.A.M.: COMPANION project MMI990002A&B, NLR reference 2005.03.31
3. Flight Crew Error / Flight Crew Performance Considerations in the Flight Deck Certification Process, Human Factors-Harmonization Working Group; Federal Aviation Administration – USA; European Aviation Safety Agency – Europe (2004)

Operator Assistance and Semi-autonomous Functions as Key Elements of Future Systems for Multiple Uav Guidance

Michael Kriegel, Claudia Meitinger, and Axel Schulte

Munich University of the German Armed Forces
Department of Aerospace Engineering (LRT-13),
85577 Neubiberg, Germany
{michael.kriegel,claudia.meitinger,axel.schulte}@unibw.de

Abstract. This paper will discuss technologies for use in a work system comprising a single operator working in a control station being supposed to supervise several UAVs. The overall setup of a typical manned-unmanned teaming scenario will be described and analyzed using the approach of the work system as a human factors engineering framework. This approach facilitates to identify the research areas of cognitive operator assistance and semi-autonomous guidance of co-operating UAVs. Furthermore, this paper will refer to recent research activities and experimental facilities for the evaluation of the solutions.

Keywords: Cognitive automation, UAV guidance, operator assistance.

1 Introduction

Future military aerial missions in asymmetric warfare or peacekeeping will be characterized by growing information demand onboard the operating units. This information demand asks for high timeliness, which can be encountered by using deposable sensor platforms for real-time and locally close reconnaissance. These platforms can be Uninhabited Aerial Vehicles (UAVs), equipped with appropriate sensors, operating in proximity of the manned mission asset, collecting information based on reconnaissance assignments. A possible scenario could comprise a manned helicopter with several UAVs being guided from a human operator onboard that helicopter. Regarding the UAV guidance there is need for the consideration and investigation of a vehicle to operator ratio larger than one. This typically leads to the demand for extensive automation. Generally spoken, traditional automation approaches provide systems, which can be regarded as tools or equipment supporting the operator to fulfill certain well defined sub-tasks. This leaves the operator in the role of the high capacity decision component determining and supervising the work process. Suchlike *conventional automation* does not assist the operator in performing tasks like decision-making with the aim to achieve the top-level mission goals because it lacks a comprehensive understanding of the situation and the overall work objective as well as decision-making and problem solving on the basis of knowledge processing.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 705–715, 2007.

In order to cope with these deficiencies, *cognitive automation* [1] shall be used as the underlying paradigm. This enables the automation to offer semi-autonomous and co-operative capabilities. Here, UAVs are enabled to accomplish missions semi-autonomously with respect to a given mission objective and to co-operate with both, other UAVs and the human operator. As part of the control station, cognitive operator assistant systems shall be enabled to assist the human in achieving the mission specified by the given mission objective and according to the situation. These operator assistant systems shall also be able to change the level of automation in order to adapt to the current workload, i.e. the availability of mental resources of the human operator.

This paper describes the overall set-up of a typical manned-unmanned teaming (MUM-T) scenario and analyses it with a top-down approach using a human factors minded systems engineering framework. It will identify and describe two relevant research areas, firstly *cognitive assistant systems* including automation capable of adapting to situational changes, and secondly *semi-autonomous* and *co-operative behavior* of UAVs facilitating multiple vehicle guidance by a single operator. Finally, relevant research project, results and experimental facilities will be described.

2 Relevant Issues Concerning Multiple Uav Guidance

By use of *conventional automation* UAVs can perform predefined tasks automatically, thereby unburdening the human operator from various sometimes tedious or even dangerous routine works. Thus uninhabited vehicles will in most cases not require a human for manual control. Although being prerequisite for safety critical performance, today there are only few and very conservative approaches on the market, concerning automatic on-board decision-making capabilities handling contingencies such as the loss of data link connection. Despite all these technological endeavors human operators will still be involved in higher level tasks such as planning, problem solving or pursuit of the overall mission goal. So, in any case we are considering a human-machine system with spatially dislocated components, as being discussed later. In fact, it is very undesirable to design fully automated systems without the possibility for a human operator to (re-)define the mission goals as well as to interact and engage during the mission process.

When regarding operation and flight guidance concepts of actual UAVs in service it can be stated that various concepts are applied. Some vehicles have to be started and landed manually by pilots standing at the runway possessing line-of-sight (e.g. *Hunter*, *Pioneer*) and beyond these flight phases control is transferred to a remote pilot in a ground control station (GCS). Other vehicles like the *Predator* will be controlled manually during takeoff and landings by use of a nose camera from the GCS, whilst during the mission phase the vehicle is controlled by an autopilot. Other vehicles (e.g. *Global Hawk*, *Shadow*) are fully automated during all flight phases (including takeoff and landing). All these flight guidance concepts require an operator to vehicle ratio of one or even larger. This is one of the factors leading to inter-crew related problems as well as fundamental flight guidance problems based on human factors implications. Studies on accidents and incidents of UAVs show that a notable percentage of mishaps are human factors related [2]. These range from display and HMI (Human-Machine Interface) problems over premature software versions to

procedural errors including wrong decision making on the part of the operator crew [3]. As stated above, conventionally designed automation fulfills strictly specified sub-tasks leaving the operator in the role of the supervisor coordinating the numerous automated functions in order to comply with the mission objectives. Within the design phase of automation systems it is difficult to anticipate all possible states and contingencies that might occur during operation. And even if the automation works as designed, unintended consequences can occur due to events that were not anticipated. Some examples are illustrated in [4]. When trying to reduce the operator to vehicle ratio to one and below (i.e. single operator guiding a single/multiple UAVs) the inter-crew related problems may be solved although the expected work load level will exceed the available human resources. Therefore assistant systems, human – machine and machine – machine collaboration are an appropriate remedy to lower the workload of the human operator. All these systems will be based on *cognitive automation* as the underlying paradigm. Further details on this approach will be portrayed in the following chapters.

3 Work System Analysis of a MUM-T Scenario

In order to be able to detail approaches for coping with problems anticipated for the guidance of multiple UAVs, this section will analyze a typical MUM-T set-up using the work system as engineering framework. Afterwards, different possibilities to introduce automation into the work system will be presented and the *Cognitive Process* as approach to the realization of artificial cognition will be explained.

3.1 The Work System as a Human Factors Engineering Framework

The work system (see figure 1) as a general ergonomics concept [5] has been utilized in a modified definition and adapted to the application domain of flight guidance by Onken [1] and UAV guidance by Schulte & Meitinger [6]. It is defined by the *work objective*, being the main input to the process of work. Usually, the work objective comes as an instruction, order or command from a supervising agency. Further constraining factors for the work process are *environmental conditions* including useful information and *supplies*. On its output the work system provides the current *state of the work* and finally a *work product* representing what has been accomplished by the work process. [7]

The work system itself consists of two major elements, i.e. the *operating force* (OF) and the *operation-supporting means* (OSMs). The *operating force* is the high-end decision component of the work system with the highest authority level. It is the only component which pursues the overall work objective. Therefore, it determines what will happen in the course of the work process and which OSMs will be deployed at what time. One major characteristic of especially a human as OF is the capability to define the work objective himself. Apart from operating on the basis of full authority competence this is the decisive criterion for what we call an *autonomous system*. The concept of the *operation-supporting means* can be seen as a container for whatever machinery or technology at the work place is available. Common to the nature of various operation-supporting means is the fact that they only perform certain well-defined sub-tasks assigned to them by the operating force. [7]

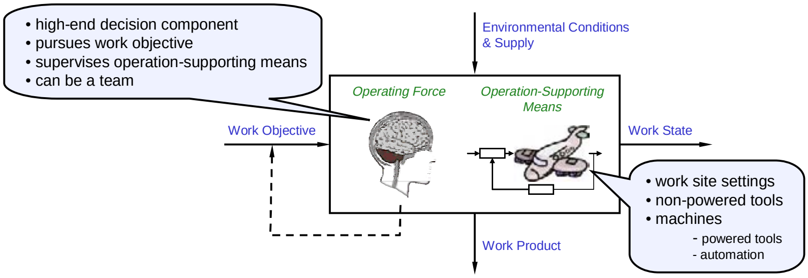


Fig. 1. The work system

In the following, the work system will be used to analyze a typical MUM-T scenario, which will serve as an example for the deduction of design criteria for introduction of advanced automation into the work system. The set-up consists of a manned helicopter being guided by a human pilot and several UAVs being guided by a human UAV operator onboard the helicopter.

Figure 2 shows one possible work system configuration consisting of two work systems – the *helicopter work system* being composed of the helicopter pilot and the helicopter itself and the *UAV guidance work system* being made up of the UAV operator and several UAVs. Here, both work systems receive the same mission objective from a superior command and control authority, while reconnaissance demands of the helicopter work system pose constraints for the UAV guidance work system. In contrast to an also imaginable configuration, in which the UAV guidance work system is subordinate to the helicopter work system, here the UAV operator can take the initiative with respect to the overall mission objective.

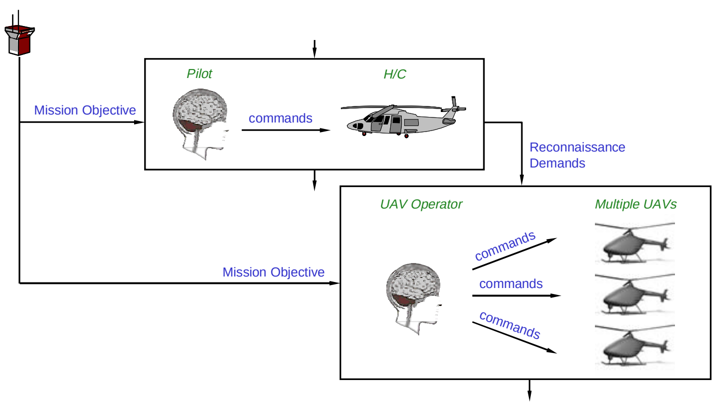


Fig. 2. System of work systems for guidance of multiple UAVs from an airborne control station

Given this arrangement, problems are likely to occur such that both humans, but especially the UAV operator, will be overloaded with their tasks, i.e. the guidance of the helicopter or the UAVs respectively. In order to keep the humans’ workload on an acceptable level, automation can be introduced into the work system in different

ways, one of which will be discussed in the next section using the UAV guidance work system as an example.

3.2 Introduction of Cognitive Automation into the UAV Guidance Work System

In order to increase the productivity and efficiency of work systems, we propose to introduce cognitive automation which promises to avoid the problems of conventional automation as sketched in section 1. Common to all cognitive automation is that it "*works on the basis of comprehensive knowledge about the work process objectives and goals […], pertinent task options and necessary data describing the current situation in the work process*" [1]. There are two ways to introduce such cognitive automation into the work system, namely as *semi-autonomous systems* or *assistant systems* (cf. [1][6][8]). *Semi-autonomous Systems* can be understood as former work systems, in which advancing automation has become capable of pursuing the given work objective and has taken over the role of the human operator as part of the operating force, but is **not** allowed to define or alter the work objective. They are always part of the operation-supporting means of a newly created work system. Thus, they are tasked by the respective operating force and capable of accomplishing these tasks in a goal-directed manner, taking the current situation into account.

The primary task of *assistant systems*, them being part of the operating force, is to support a human operator in pursuing the given work objective. Thus, the main difference to semi-autonomous systems is the requirement to know about, understand and pursue the work objective in contrast to assigned tasks. In order to be able to support the human adapted to his or her current needs, an assistant system moreover has to understand human resources and be capable of human-machine co-operation.

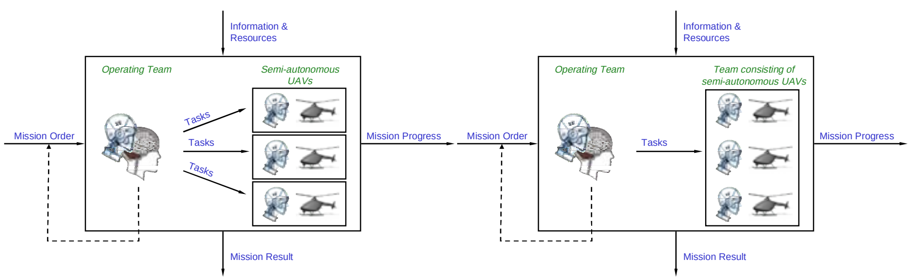


Fig. 3. Introduction of ACUs into the UAV guidance work system

Figure 3 (left) shows an obvious approach to introduce semi-autonomous and assistant systems into the UAV guidance work system as introduced in the previous section. In a first step, each UAV is equipped with an *Artificial Cognitive Unit (ACU)* (depicted by a robot head) being capable to accomplish tasks as opposed to the execution of detailed instructions, thus, each forming a semi-autonomous system. In order to be able to cope with his or her primary task, i.e. to supervise several semi-autonomous systems and allocate tasks to them in order to achieve the work objective, the human operator will be assisted by an ACU. A further improvement of the operator-to-UAV ratio may be achieved by enabling the semi-autonomous UAVs to

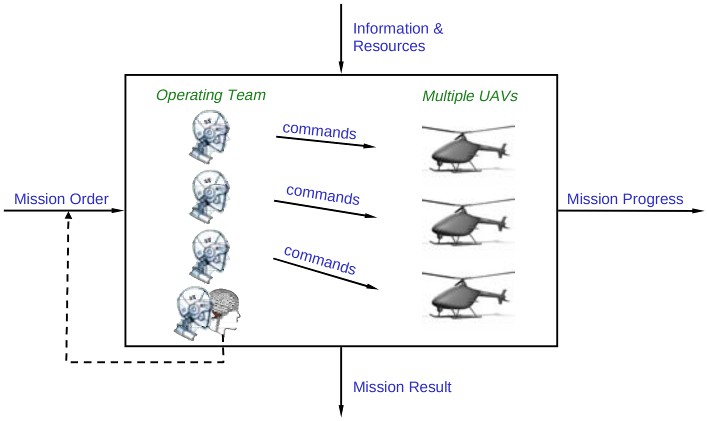


Fig. 4. Alternative Introduction of ACUs into the UAV guidance work system

co-operate (see Figure 3, right). Here, the UAVs can be tasked as a team, thus taking over the co-ordination task from the operating force.

Figure 4 shows an alternative configuration acting on the maxim to put as much automation as possible into the OF (cf. [7]). This recommendation is based on the assumption that the more automation knows about the work objective to achieve, the better decisions can be made. Here, the ACUs being responsible for the guidance of the UAVs become part of the OF, thus pursuing the given work objective and acting in co-operation with the human. Such teamwork also increases the involvement of the human operator as opposed to a mere supervisory participation in the work process.

3.3 The Cognitive Process

As indicated above, both semi-autonomous and assistant systems can come in the shape of *Artificial Cognitive Units (ACUs)* being capable of goal-directed and rational behavior. As underlying theory for the realization of such systems, the *Cognitive Process* is suggested, which is an adequate model of human information processing [9].

It follows a knowledge-based approach, i.e. separates knowledge from knowledge processing. The behavior of systems developed according to the theory of the Cognitive Process is mainly driven by the *a-priori knowledge*, which is modeled by the system developer. There are four kinds of a-priori knowledge. *Environment models* represent objects, relations, and abstract concepts, which are expected to occur in the environment or be relevant for the mission in some way. They are used to gather a belief about the situation of the environment. *Desires* model potential goals, at which the central question is, which goals should be pursued. The set of active goals represents a situation, which shall be achieved. *Action alternatives* are used to put together a plan, which is suited to transfer the current situation into the desired one. Finally, *instruction models* are used to determine the instructions which can execute the elements of the plan and finally effects and change the environment.

The paradigm of the Cognitive Process can be used to develop different application capabilities separately, the knowledge belonging to each capability being encapsulated in so-called *packages*. These packages are linked by dedicated joints in the a-priori knowledge and together form the complete system.

4 Cognitive UAV Co-operation

Having introduced a team of semi-autonomous UAVs into the UAV guidance work system and presented the *Cognitive Process* as underlying theory for realization of artificial cognition, this section details an approach to implement co-operative capabilities of semi-autonomous systems, which provides a basis for human-machine co-operation as discussed in the following section.

In general, *co-operation* is characterized by a common objective, which is pursued by all members of a team being committed to its achievement. In the context of the work system, the common objective is usually the work objective for a team in the operating force and an assigned task for a team being part of the operation-supporting means. In order to achieve the common objective, team members have to *coordinate*, i.e. manage dependencies among their activities [10]. This includes task allocation within the team as well as the assignment of shared resources or the temporal arrangement of tasks, the execution of which e.g. depends on the successful completion of another task. Coordination among team members in turn requires *communication*. In this context we distinguish between explicit and implicit communication. Explicit communication includes the exchange of messages as well as non-verbal actions, which are executed in order to provide some information to somebody else. Implicit communication in contrast deduces information by observation of actions, the primary purpose of which is not to communicate, but to e.g. perform a mission-related task.

In order to facilitate co-operative behavior of ACUs, the appropriate models of the a-priori knowledge within the Cognitive Process have to be developed. According to the *Cognitive Process Method* [11], at first, desires have to be modeled, as they finally drive the behavior. Top-level desires being relevant in the context of co-operative behavior are based on requirements for human-machine co-operation by [12] and training of human teams [13] and cover teamwork per se as well as the achievement of the common objective and coordination and communication aspects [14]. Action alternatives which are suited to achieve co-operation goals are modeled next and either refer to information exchange or the assignment of tasks or resources to team members.

A more detailed description of the a-priori knowledge necessary for the realization of co-operative ACU behavior is given by [14]. This paper also describes the results obtained with an implementation of such capabilities taking a simplified Sead-/Attack mission performed by five UAVs with heterogeneous capabilities as an example.

Finally, it shall be emphasized, that successful co-operation of several semi-autonomous systems always depends on the capabilities of individual team members to actually comply with the responsibilities which have been assigned to them.

5 Cognitive and Co-operative Operator Assistance

Another research area to be addressed here is the development of cognitively co-operating UAV operator assistant systems. For such operator assistant systems the basic requirements for pilot assistant systems stated by Onken [15] can also be applied. They state that the attention of the operator has to be directed to the most important task in the current situation and, if this has been achieved, and there is still a situation with human overload, this situation has to be transferred into a normal one by the use of technical means.

In order to be able to realize such functionalities the *Cognitive Process* (see section 3.3) shall be used as the underlying paradigm for the implementation. Concerning a UAV operator assistant system, mainly two packages will be required. The first one shall address the domain of UAV guidance and could thus be called "*domain expert*". Environment models of this package incorporate knowledge about the work objectives (usually a mission order), environmental elements and conditions, and the UAVs including their states and capabilities. The desires include models such as to accomplish the mission and consider the tactical situation. Possible action alternatives cover the deployment of available operation-supporting means i.e. the UAVs and their on-board automation functions, depending on their resources and capabilities.

The second package shall address the human operator and could thus be called "*operator assistance*". Environment models for this package include models to evaluate resources and behaviors of the human operator as well as his or her intents and possible errors. The desires incorporate models to form a team with the human operator and to jointly achieve the common work objective (cf. section 4). Action alternatives cover models regarding interaction and dialog management with the human operator. Forthcoming works in the context of our MUM-T project (see next section) are aiming for the realization and evaluation of such a system.

6 Projects and Experimental Facilities

Two already completed projects, namely CAMA (Crew Assistance Military Aircraft) and CASSY (Cockpit ASsitant SYstem) have been accomplished and tested under real world conditions in flight campaigns [16][17][18]. Main focus of research in these projects were the development of a fully automatic on-board *flight planner* and a *pilot model* used to anticipate expected pilots' actions depending on the mission phase. On the basis of this, pilot errors are distinguished from intended but deviating pilot actions. In any case, the assistant system will adapt its interventions accordingly. Ongoing projects include PILAS (Pilot Assistant System) where an assistant system is designed to track the actual mission phase and to support the operator depending on the current situation if appropriate [19]. Another ongoing project is MiRA (Military Rotorcraft Associate) where the focal point is set on adaptive function allocation for operator workload adjustment. The research on machine-machine co-operation is the main subject to the project COSY*team*, where several UAVs are accomplishing a given mission in co-operation under loose, so far unassisted human supervision [14].

The last project to be mentioned here is MUM-T (manned- unmanned teaming). It combines all aforementioned research areas, i.e. multiple UAV guidance from an

airborne platform. Therefore, a surrogate scenario is set up by keeping the same work system relationships as explained earlier in this paper. This experimental system is called Co²SiMA (Cognitive & Co-operative System for intelligent Mission Accomplishment). Co²SiMA facilitates a mobile control station on the basis of a Mercedes Sprinter truck. This component will be the surrogate for the manned helicopter with full functionality of the UAV-operator station. The second main component of Co²SiMA is comprised of a fleet of flying model based type-different UAVs (rotorcraft and fixed-wing aerial vehicles) equipped with adequate technology (e.g. autopilot, flight management system, hardware infrastructure to host the cognitive functionalities, sensors e.g. AHRS, DGPS, CCTV and data links). Figure 5 depicts the main components of Co²SiMA.

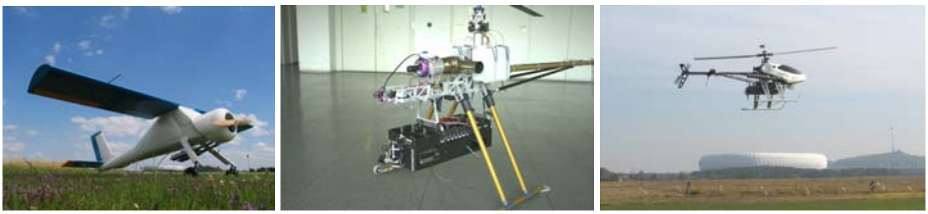


Fig. 5. Main components of Co²SiMA

The surrogate scenario envisions the mobile control station being guided by the reconnaissance information (e.g. real-time TV stream) provided by the UAVs through the UAV operator. This setup allows the testing and evaluation of developed technologies under simulated and real world conditions. Experimental focus will be set on validating the performance of the UAV-UAV co-operation, the investigation of a operator to vehicle ratio smaller than one, the validation and possible application of operator models as well as the appliance of human factors related assessments and measurement techniques (e.g. eye movements, workload). Future developments of co-operative assistant systems will benefit from these experimental resources and results. Concerning real-world trials the focal point will be set on the possible viability of critical functionalities and capabilities.

7 Conclusion

Manned-unmanned teaming, as it is understood within our research group, poses new and heavily demanding task load on operators. On the one hand there should be mentioned the adverse work environment of an airborne platform maneuvering in a threatened military theatre. On the other hand there is the demand for interaction with necessarily highly automated systems, i.e. multiple UAVs designated to accomplish a complex mission in a coordinated manner. This entails both, high mental workload as well as extreme load on attention allocation and situation awareness processes. The proposed solution, as recommended by our research group and presented in this paper, is the approach of cognitive and co-operative automation. Various recent and

current research activities in the fields of knowledge-based operator assistant systems and co-operative semi-autonomous UAV flight guidance systems, which led to very promising results, point out the way ahead. Within a German MoD funded research project these findings shall be transferred to the domain of manned-unmanned military helicopter missions. This includes the development of a UAV-operator assistant system and semi-autonomous co-operating UAVs as remote sensor platforms.

References

1. Onken, R.: Cognitive Cooperation for the Sake of the Human-Machine Team Effectiveness. In: RTO-Meeting Proceedings MP-088 (HFM-084). Warsaw, Poland (October 7-9, 2002)
2. Williams, K.W.: A Summary of Unmanned Aircraft Accident/Incident Data: Human Factors Implications. FAA Civil Aerospace Medical Institute, Oklahoma City, OK, USA (2004)
3. Tvaryanas, A.P., Thompson, W.T., Constable, S.H.: US Military Unmanned Aerial Vehicle Mishaps: Assessment of the Role of Human Factors Using HFACS (2005)
4. Williams, K.W.: Human Factors Implications of Unmanned Aircraft Accidents: Flight Control Problems. FAA Civil Aerospace Medical Institute, Oklahoma City, OK, USA (2006)
5. REFA. Methodenlehre des Arbeitsstudiums, Teil 1-3. 7. Auflage. Verband für Arbeitsstudien und Betriebsorganisation. München: Carl Hanser Verlag (1984)
6. Schulte, A., Meitinger, C.: Human Factors in the Guidance of Autonomous Vehicles: Oxymoron or Tautology? The Potential of Cognitive and Co-operative Operator Assistant Systems (To appear)
7. Onken, R., Schulte, A.: Engineering of Cognitive Systems for Vehicle Guidance and Control Work Processes (To appear)
8. Kriegel, M., Schulte, A.: Work System Analysis of the Integration of Autonomous Functions and Intelligent Operator Assistance in UAV Guidance. In: RTO-Meeting Proceedings HFM-135. Biarritz, France (October 9-11, 2006)
9. Putzer, H., Onken, R.: COSA – A generic cognitive system architecture based on a cognitive model of human behaviour. International Journal on Cognition, Technology and Work. 5, 140–151 (2003)
10. Malone, T.W., Crowston, K.: The Interdisciplinary Study of Coordination. ACM Computing Surveys 26(1), 87–119 (1994)
11. Putzer, H.: Ein uniformer Architekturansatz für kogitive Systeme und seine Umsetzung in ein operatives Framework. Berlin: Köster (2004)
12. Billings, C.E.: Aviation Automation: The Search for a Human-Centred Approach. Lawrence Erlbaum Associates, Mahwah (1997)
13. Swezey, R.W., Salas, E.: Guidelines for Use in Team-Training Development. In: Swezey, R.W., Salas, E. (eds.) Teams: Their Training and Performance, Ablex Publishing Corporation, Norwood, NJ (1992)
14. Meitinger, C., Schulte, A.: Cognitive Machine Co-operation as Basis for Guidance of Multiple UAVs. In: RTO-Meeting Proceedings HFM-135. Biarritz, France (2006)
15. Onken, R.: Basic Requirements Concerning Man-Machine Interactions in Combat Aircraft. In: Workshop on Human Factors / Future Combat Aircraft, Ottobrunn, Germany (1994)

16. Prévôt, T., Gerlach, M., Ruckdeschel, W., Wittig, T., Onken, R.: Evaluation of intelligent on-board pilot assistance in in-flight field trials. In: 6th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems. Massachusetts Institute of Technology. Cambridge, MA. (June 1995)
17. Stütz, P., Schulte, A.: Evaluation of the Cockpit Assistant Military Aircraft CAMA in flight trials. In: Third International Conference on Engineering Psychology and Cognitive Ergonomics. Edinburgh, UK. (October 25-27, 2000)
18. Frey, A., Lenz, A., Putzer, H., Walsdorf, A., Onken, R.: In-Flight Evaluation of CAMA – The Crew Assistant Military Aircraft. In: Deutscher Luft- und Raumfahrtkongress. Hamburg, Germany (September 17-20, 2001)
19. Groth, C., Meitinger, C., Donath, D., Schulte, A.: Missionsauftragsanalyse in COSA als Funktionsmodule eines Pilotenassistenzsystems. In: Deutscher Luft- und Raumfahrtkongress. Braunschweig, Germany (2006)

Confucius in Western Cockpits: The Investigation of Long-Term Versus Short-Term Orientation Culture and Aviation Accidents

Wen-Chin Li and Don Harris

Department of Human Factors, Cranfield University,
Bedfordshire, United Kingdom
{w.li.2002,d.harris}@cranfield.ac.uk

Abstract. This research applies the Human Factors Analysis and Classification System and Hofsted's fifth dimension of national culture (Confucian- long-term versus short-term orientation) to compare accident patterns between the US and Taiwan. Asia and Africa have higher accident rates than Europe or America. There are also fundamental differences between Chinese and Western minds. These variations suggest that there should be fundamental, underlying factors causing these differences. Several studies have investigated the relationship between culture and accidents however, no research has investigated Chinese culture and accidents. The findings clearly show different patterns in the human factors causes underlying aviation accidents in these different regions. It could even be argued that the accident analysis system itself has an implicit cultural bias within it, as HFACS was a product of Western culture. Global aviation is strongly influenced by the Western culture, however, the safety challenge is to manage the potential risks it may present.

Keywords: Accident Investigation, Confucian, Cross-culture, Human Error, Human Factors Analysis and Classification System.

1 Introduction

It is generally acknowledged that the aviation accident rates differ across regions. Asia and Africa have higher accident rates than either Europe or America. These regional variations suggest that there were fundamental, underlying factors causing these differences in accident rates. Furthermore, it can be suggested that the majority of the facets of the aviation system has been constructed from a Western (North American/Western European) perspective (Klein, 2004). As a result, the causal factors underlying accidents and prevention strategies that seem reasonable to Westerners might present problems for East Asian and African people. What is more, Western people might not even be aware of such a problem (Jing, Lu & Peng, 2001). There has been a great deal of debate about the role of culture in aviation mishaps, however, culture is rarely cited as a causal factor underlying accidents. Nevertheless, culture is

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 716–722, 2007.

at the root of action; it underlies the manner by which people communicate and develop attitudes towards life. There are many definitions of culture. A culture is formed by its environment and evolves in response to changes in that environment, therefore, culture and context are really inseparable (Merritt & Maurino, 2004).

This research applied the Human Factors Analysis and Classification System (Wiegmann & Shappell, 2003) and Hofsted's (2001) fifth dimension of national culture, Confucian- long-term versus short-term orientation to compare the accident pattern between the US and Taiwan. Confucius's teachings are lessons in practical ethics. People had to rely upon themselves, not some external force, to maintain their humanity. The following are key principles of Confucian teaching, (1) the stability of society is based on unequal relationships between people. These relationship are based on mutual and complementary obligations. Junior owes the senior respect and obedience; the senior owes the junior protection and consideration; (2) the family is the prototype of all social organization. A person is not primarily and individual, harmony is found in the maintenance of everybody's face, in the sense of dignity, self-respect and prestige; (3) virtuous behaviour toward others consists of not treating others as one would not like to be treated oneself; (4) virtue with regard to one's tasks in life consists of trying to acquire skills and education, working hard, not spending more than necessary, being patient, and persevering, conspicuous consumption is taboo, as is losing one's temper. There are fundamental difference between Chinese minds and Western. In science and technology, Western Truth stimulated analytic thinking, whereas Eastern Virtue led to synthetic thinking (Hofstede, 2001). Through their different logics East and West followed different paths in developing government and in developing science and technology. Whereas the Romans spread the principle of 'government by law', the main continuous principle of Chinese was 'government by man'.

Culture fashions a complex framework of national, organizational and professional attitudes and values within which groups and individuals function. Cultures can be divided into different levels: families, organizations, professions, regions, and countries. The power of culture often goes unrecognized since it represents 'the way we do things here'. It is the natural and unquestioned mode of viewing the world and national cultural characteristics play a significant part in aviation safety (Helmreich & Merritt, 1998). There were several studies investigated the relationship of culture and accident pattern (Helmreich & Merritt, 1998; Johnston, 1993; Merritt & Maurino, 2004; Orasanu & Connolly, 1993; Soeters & Boer. 2000). However, no research illustrated Chinese culture and aviation accident's pattern. There is a raising need for investigating the relationship between Chinese culture and safety of aviation operation, as the both of Chinese population and market for aviation industry. To summarize, there is an ongoing need to understand the influence of culture on aviation safety. Detailed examination of the relative incidence of the underlying human factors components in the causation of accidents using HFACS will provide greater insight in this respect, supplementing and adding explanatory power to the observation that accident rates differ between different countries and cultures. This research examines

the relative frequency of contributory factors using the HFACS framework from aircraft accidents in Taiwan (Chinese culture) and the USA, and relates these differences to aspects of regional culture as described using the typology proposed by Hofstede (1991 & 2001).

2 Method

Data: The data analyzed in the present study are taken from two researches classifying aviation mishaps using the HFACS framework. These are from Taiwan and the USA. There were 523 accidents with 1,762 instances of human error categorized using the HFACS framework from data elicited from the Taiwan Air Force between 1978 and 2002 (Li & Harris, 2005); and 119 accidents with 319 of categorized instanced of human error in US data recorded between 1990 and 1996 (Wiegmann & Shappell, 2001). According to Hofstede's (1991 & 2001) fifth culture dimension, Confucian (long-term versus short-term orientation), the score of Taiwan is 87, the score of US is 29, the world average is 45. It is clear that Taiwan is long-term orientation culture, US is short-term orientation culture. It is hypothesized that different cultures will show different patterns in the underlying causal factors in aircraft accidents.

Classification framework: This study based on the HFACS framework as described in Wiegmann & Shappell (2003). The first level of HFACS categorizes is 'unsafe acts of operators' that can lead to an accident including and comprises of four subcategories of 'decision errors'; 'skill-based errors'; 'perceptual errors' and 'violations' . The second level of HFACS concerns 'preconditions for unsafe acts' which has a further seven sub-categories of 'adverse mental states'; 'adverse physiological states'; 'physical/mental limitations'; 'crew resource management'; 'personal readiness'; 'physical environment', and 'technological environment'. The third level of HFACS is 'unsafe supervision' including 'inadequate supervision'; 'planned inappropriate operation'; 'failure to correct known problem', and 'supervisory violation'. The fourth and highest level of HFACS is 'organizational influences' and comprises of the subcategories of 'resource management'; 'organizational climate' and 'organizational process'.

To avoid over-representation from any single accident, each HFACS category was counted a maximum of only once per accident. These counts acted simply as an indicator of presence or absence of each of the 18 categories in any given accident. These data were then subject to chi-square (χ^2) analyses to measure the statistical strength of association between HFACS category and country.

Reliability of HFACS Framework: Inter-rater reliabilities of the data from Taiwan, calculated as a simple percentage rate of agreement, obtained reliability figures for the 18 categories of HFACS of between 72.3% and 96.4% (2005). The average of the inter-rater reliabilities of the data gathered from the US data showed 76% (2001).

Table 1. The Frequency of HFACS Categories Between Taiwan and USA

HFACS Categories	Yes	No	Ye	No	Chi-square (χ^2)and Goodman & Kruskal Tau (τ)
Level-4, Organizational process	**76** ***80***	447 *443*	**10** ***18***	109 *100*	χ^2= 3.13, df=1, p=.076
Organizational climate	**4** ***7***	519 *516*	**0** ***2***	119 *117*	χ^2= 0.91, df=1, p=.338
Resource management	184 *156*	339 *366*	**3** ***36***	116 *83*	χ^2= 50.09, df=1, p=.000
Level-3, Supervisory violation	8 *9*	515 *514*	2 *2*	117 *117*	χ^2= 0.01, df=1, p=.904
Failed correct a known problem	12 *12*	511 *511*	2 *3*	117 *116*	χ^2= 0.17, df=1, p=.679
Planned inadequate operations	24 *22*	499 *501*	1 *5*	118 *114*	χ^2= 3.64, df=1, p=.056
Inadequate supervision	177 *144*	346 *379*	**6** ***33***	113 *86*	χ^2= 39.45, df=1, p=.000
Level-2, Technology environment	44 na	479 na	na na	na na	na
Physical environment	74 na	449 na	na na	na na	na
Personal readiness	29 *25*	494 *498*	**0** ***6***	119 *113*	χ^2= 6.91, df=1, p=.008
Crew resource management	146 *142*	377 *381*	35 *32*	84 *87*	χ^2=0.10, df= 2, p=.743
Physical/mental limitation	**73** ***77***	450 *446*	**13** ***17***	106 *102*	χ^2= 0.76, df=1, p=.380
Adverse physiological states	**2** ***5***	521 *518*	2 *1*	117 *118*	χ^2=2.63, df=1, p=.104
Adverse mental states	184 *156*	339 *367*	**16** ***36***	103 *83*	χ^2= 21.35, df=1, p=.000
Level-1, Violations	160 *158*	363 *365*	32 *36*	87 *83*	χ^2= 0.63, df=1, p=.426
Perceptual errors	116 *106*	407 *417*	17 *24*	102 *95*	χ^2= 3.67, df=1, p=.055
Skilled-based errors	**226** ***245***	297 *278*	72 56	47 *63*	χ^2= 11.65, df=1, p=.000
Decision errors	223 *202*	300 *321*	**34** ***46***	85 *73*	χ^2= 7.99, df=1, p=.004

Note: 1. Numbers show as a Roman font at the top of each cell is the observed value; the numbers shown in an Italic font at the bottom of each cell is the expected value.

2. 'na' indicates no information was available for the categories of 'technology environment' and 'physical environment'.

3. Bold font indicates under-representative.

4. Big size font indicates over-representative.

3 Results

There were six HFACS categories exhibited significant differences in reported frequency of aviation accidents between Taiwan and US (Table 1). These were, ‘resource management’ (level 4); ‘inadequate supervision’ (level 3); ‘personal readiness’ and ‘adverse mental states’ (level 2); ‘skilled-based errors’ and ‘decision errors’ (level 1). Furthermore ‘organizational process’ (level 4); ‘planned inadequate operations’ (level 3) and ‘perceptual errors’ (level 1) were verging on statistical significance ($p<0.10$).

Level 4 - Organizational Influences: There was one HFACS category with significant difference between Taiwan and US (table 1). ‘Resource management’, which includes the selection, staffing and training of human resources at an organizational level, excessive cost cutting, providing unsuitable equipment, and a failure to remedy design flaws, was over-represented in Taiwan and was under-represented in US.

Level 3 - Unsafe Supervision: There was one HFACS category which exhibited significant differences in recorded frequency of being implicated in accidents (table 1). This was ‘inadequate supervision’ which includes factors such as a failure to provide proper training, adequate rest periods, a lack of accountability, failure to track qualifications and performance, using untrained supervisors and loss of situation awareness at the supervisory level. This category was over-represented in the Taiwan sample and under-represented in the USA.

Level 2 - Preconditions for Unsafe Acts: There were two categories with significant differences in frequency of occurrence between Taiwan and US (table 1). ’Adverse mental states’, which includes issues such as over-confidence, stress, loss of situational awareness, distraction, channelized attention and task saturation, was over-represented in the Taiwan sample, and under-represented in the USA. ‘Personal readiness’ which encompassed issues associated with inadequate training, self-medication, poor diet, and overexertion while off duty, was over-represented in frequency of occurrence in Taiwan, and under-represented in US accidents.

Level 1 - Unsafe Acts of Operators: There were two HFACS categories which showed differences in their frequency of occurrence between regions at level-1 (table 1). ‘Skill-based errors’ which includes actions such as inappropriate stick and rudder coordination, excessive use of flight controls, glide path not maintained, and adopting an improper airspeed or altitude, was over-represent in US accidents but under-represented in the sample from Taiwan. ‘Decision errors’, which includes issues such as selecting inappropriate strategies to perform a mission, improper in-flight planning, making an inappropriate decision to abort a take-off or landing or using improper remedial actions in an emergency, was over-represented in Taiwan sample and under-represented in the US.

4 Discussion

When Western (North American/Western European) engineers and human factor specialists develop equipment, training and procedures, they incorporate their own vision of the world which is heavily influenced by the cultural norms of their country.

They implicitly assume that all users around the world will share their reasoning and values. Klein (2004) observed that people from different nations differ in their cognition in ways that result in dissimilar perceptions, judgments and decision-making. National culture provides a fundamental basis for a group member's behavior, social roles and cognitive processes. It also provides underlying rules about safety, effective communication, and provides the basis for verbal and nonverbal interactions. This research, using the HFACS framework suggests that there are statistically significant differences in the relative frequencies of the underlying human factors causes in aviation mishap between Taiwan and US. However, such a simple analysis alone showing differences between regions has little explanatory power. It is essential to identify the potential causal roots for these differences in relative frequency of the underlying factors in these aviation mishaps.

According to Hofstede’s (2001) fifth dimension of national cultures, long-term versus short-term orientation, was found to be based on items reminiscent of the teachings of Confucius, on both of its poles: persistence and thrift to personal stability and respect for tradition. The US culture has strong desire searching for truth and governed by ‘law’. The Chinese tradition does not hold laws and abstract principles in high regard and governed by ‘man’. This could possibly explain the Taiwan with higher accident rate in ‘resource management’ (level-4) than US. Furthermore, Western culture believed in absolute guidelines about good and evil, Chinese culture believed what is good and evil depends on the circumstances. It might illustrated the US data has lower accident rate at the category of ‘inadequate supervision’ (level-3) than Taiwan. The supervisory levels in Taiwan were not following strict principles to perform their duties caused problems. The cultural difference of ‘probabilistic thinking of Western’ with ‘either full or no confidence of Chinese’ might describe the US with lower ‘adverse mental states’ (level-2) accidents than Taiwan. The attitude of overconfidence or no confidence in long-tern orientation culture was not a safe condition for conducting flight operations. Also, the ‘perseverance’ in long-term orientation culture undermined precondition for unsafe acts and caused higher ‘personal readiness’ problem (level-2) in Taiwan than US.

Accidents in the US sample are only over-represented at level-1 ‘unsafe acts of operators’ (with over-representation in the ‘skilled-based errors’ category). It may be suggested that the explanation for these observations is that the US has a culture which prefers individual decision making and responsibility for the self. In Hofstede’s (1991) terms it is an ego-oriented society. Chinese place less value on 'cognitive consistency'. Also, it has been shown that in comparison with the US, Chinese view 'disagreement' as less face-threatening to personal relationships than 'injury' or 'disappointment'. A different opinion does not affect their egos so much. Through their different logics East and West followed different paths in developing government and in developing science and technology. In science and technology, Western Truth stimulated analytic thinking, whereas Eastern Virtue led to synthetic thinking (Hofstede, 2001). This cultural characteristic illustrated Taiwan with higher ‘decision errors’ (level-1) than the US. The analytic thinking approach is a safer approach than synthetic thinking in aviation domain.

5 Conclusion

The findings clearly show different patterns in the human factors causes underlying aviation accidents in different regions. The underlying cultural causes of these differences are also postulated. It should be noted, however, that the USA and Taiwan are only exemplars of the cultures that they represent. Generally the short-term orientation culture seems to be superior for promoting aviation safety compared to the long-term orientation cultures. However factors such as the design of the aircraft, the management procedures and the nature of safety regulation all have a strong Western influence of short-term orientation culture. All of these factors are culturally congruent with the USA, so it is perhaps not too surprising that this country comes out best when using the HFACS to analyze the underlying causes of accidents. It could even be argued that the accident analysis system itself has an implicit cultural bias within it, as HFACS also created by Western culture. Global aviation is strongly influenced by the Western culture, however, the challenge for safety is not to ignore these cross-cultural issues influencing safety but to manage the potential risks they may present.

References

1. Helmreich, R.L., Merritt, A.C.: Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences, Ashgate, Aldershot, England (1998)
2. Hofstede, G.: Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations. Sage Publications, California (2001)
3. Hofstede, G.: Cultures and Organizations: Software of the Mind. McGraw-Hill, London (1991)
4. Jing, H.S., Lu, C.J., Peng, S.J.: Culture, Authoritarianism and Commercial Aircraft Accident. Human Factors and Aerospace Safety 1(4), 341–359 (2001)
5. Johnston, N.: Regional and Cross-Cultural Aspects of CRM. In: Hayward, B.J., Lowe, A.R. (eds.) Australian Aviation Psychology Symposium, Sydney (1993)
6. Klein, H.A.: Cultural differences in cognition: Barriers in multinational collaborations. In: Montgomery, H.E., Lipshitz, R.E. (eds.) How professionals make decisions, pp. 243–253. Lawrence Erlbaum, Mahwah (2005)
7. Li, W.C., Harris, D.: HFACS Analysis of ROC Air Force Aviation Accidents: reliability analysis and cross-cultural comparison. International Journal of Applied Aviation Studies 5(1) (2005)
8. Merritt, A., Maurino, D.: Cross-cultural factors in aviation safety. In: Kaplan, M. (ed.) Advances in human performance and cognitive engineering research, pp. 147–181. Elsevier, Amsterdam (2004)
9. Orasanu, J., Connolly, T.: The Reinvention of Decision Making. In: Klein, G.A., Orasanu, J., Calderwood, R., Zsambok, C.E. (eds.) Decision Making in Action: Models and Methods, pp. 3–20. Ablex, Norwood, New Jersey (1993)
10. Soeters, J.L., Boer, P.C.: Culture and flight safety in military aviation. The. International Journal of Aviation Psychology 10(2), 111–133 (2000)
11. Wiegmann, D.A., Shappell, S.A.: Human Error Analysis of Commerical Aviation Accidents: Application of the Human Factors Analysis and Classification System. Aviation, Space, and Environmental Medicine 72(11), 1006–1016 (2001c)
12. Wiegmann, D.A., Shappell, S.A.: A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System, Ashgate, Aldershot, England (2003)

Voice Alarm System in Emergency Evacuation

Huiyang Li[1,2], Xianghong Sun[1], and Kan Zhang[1]

[1] Institute of psychology, Chinese Academy of Sciences
100101 Beijing, P.R. China
[2] Graduate School of Chinese Academy of Sciences
100049 Beijing, P.R. China
{lihy,sunxh,zhangk}@psych.ac.cn

Abstract. Under emergency situations such as large fires, floods, hazardous-materials, etc., incident commander have to manage an evacuation in help of alarm systems (audible or/and visual notification). This article reviewed selected literature relevant to ergonomics of alarm systems (esp. voice alarm system) in emergency evacuation, and occupants' response behavior to the voice alarm. The literature cited is of world-wide origin, and is mainly from China, Canada and U.S.. At the end of the article, future directions in the research area are recommended.

Keywords: voice alarm, alarm system, emergency evacuation.

1 Introduction

There are a variety of emergency situations that could justify an incident commander (IC) having to order an evacuation. The scenarios include large fires, floods, tornadoes and blizzards; hazardous-materials incidents (including biological, chemical, explosive, and radiological events); and civil disturbances, to name a few. Perhaps one of the rarest and most challenging elements of a major incident is managing the evacuation of a populated area. This is especially true when the affected area is a public one which has a large crowd of occupants with different backgrounds (e.g. age, gender, experience of evacuation training, common or handicapped).

Studies on human behavior in evacuation are mainly included in those investigating human behavior in fire. Researchers in safety engineering, ergonomics and psycho-logy began studying it as early as 1900. John L. Bryan defined the three periods of the study area of research on human behavior in fire: (1) The "Prerecognition Years" period which was arbitrarily selected as the period from the evacuation studies of the early 1900s to the 1970s; (2) The "Productive Years" was selected to include the 1970s and the 1980s; (3) The "Performance Code Incentive Years" were identified as the 1990s into the 2000s [1].

The main methods in those researches include: interview, questionnaire [2, 3], analyzing videotapes [4, 5], observation [6, 7], experiments [8, 9], etc.. Issues in this study area embrace: evacuation time [5], panic behavior [10], pre-evacuation behavior, behavior differences between occupants of various ages, gender, roles, evacuating experiences and relative training [5], occupants' response to the fire alarm or public

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 723–730, 2007.

address system [4, 6-9, 11, 27-29], etc.. The results and conclusions of the studies gave valuable recommendation to the IC who manage the evacuation and provided quantitive and precise parameter value references to simulation of evacuation process.

Under emergency situations, environmental factors are featured in smoke, power cut, crowd, noise, etc. and occupants in a building or a location are very likely to be panic and irrational. Studies and observation found that, in order to deliver the information fast and accurately, and make the evacuation efficient and safe, the IC should use speech signals as the alarm and instructing signal instead of using bell alarm only [11, 30]. There are some studies on speech signals and voice alarm system [12-13].

This article reviewed selected literature relevant to alarm systems (esp. voice alarm system) in emergency evacuation, and occupants' response behavior to the voice alarm. The literature cited is of world-wide origin, and is mainly from China [12-14], Canada [3-5] and U.S. [2].

2 Alarm Systems in Emergency Evacuation

During an evacuation, the purpose of the alarm system is to encourage the occupants to evacuate correctly and as quickly as possible. The amount of information that the system can impart relies on the style of alarm used and the occupant's interpretation of the alarm itself.

The effectiveness of alarm systems is fundamental to the success of an evacuation. An effective alarm system may not only reduce the time it takes for an individual to react to the emergency, but might initialize a predetermined chain of events which leads to the safe evacuation as well. The factors involved in the alarm system which effects this procedure can be summarized as: (1) The clarity of the warning; (2) The believability of the alarm [15].

The clarity of an alarm system concerns the information that the system provides to the population, and whether it clearly indicates the occurrence of an emergency incident; i.e., it is possible for the occupant population to determine the enclosure alarm from other adjacent alarm systems [16]. The level of effectiveness of alarm systems and occupants' understanding might be improved through the introduction of an IFWS (information warning system) alarm system, which includes a graphical/aural explanation of the event [16-20], or through public address systems and pre-recorded messages [21].

The sound of a voice-even a recorded voice-giving directions is considered to convey more authority than a simple bell. However, in using such systems operators must ensure that correct message is broadcast. The importance of this was tragically demonstrated in the Dusseldorf airport fire where incorrect messages were broadcast which directed occupants to the seat of the fire rather than to safety [21].

In addition to the clarity of information supplied, the clarity of the alarm system is also important to the reception of the alarm signal. This can be affected, especially in traditional bell alarms, by the location and power of the alarm signal. Members of the population who do not clearly receive alarm information may misinterpret or ignore the message [22].

The believability of the alarm is dependent on the frequency with which the system is tested on the enclosure population, the frequency of malfunctions, and the frequency

of false alarms, i.e. the accuracy of the alarms [16]. The frequency of these events significantly effects the manner in which occupants respond to alarm signals [23].

Another factor that has significant influence on the success of evacuation is the level of education in the expected actions and responses to the activity of the alarm. So it seems to go along with the effectiveness of the alarm system. The interpretation and reaction of occupants to the alarm might be improved by implementing modern developments and by using a combination of audible and visual notification instead of using a single one (either audible or visual). It will increase occupant response as well as improve the system to react intelligently to the surrounding conditions.

3 Voice Alarm System

Generally speaking, in conventional human-machine/computer systems, the human operator's auditory channel has been primarily used for transmitting verbal communication from other operators (e.g., messages to the pilot from air traffic control) and for presenting auditory warning signals (tones, horns, buzzers, etc.). In the recent ten years, however, rapid advances in microcomputer technology have produced highly efficient speech-synthesis units generating, for example, telephone menus or simple feedback to users of computer or other systems [24].

Alarms composed of synthetic voice or recorded voice message provide one answer to the problems of discriminability and confusion. Unlike "symbolic" sounds, the hearer does not need to depend on an arbitrary learned connection to associated sound with meaning. The loud sounds *"Engine fire!"* in the cockpit or "*There is fire on the 3rd floor!*" in a building mean exactly what they seem to mean. Voice alarms are employed in several circumstances. But voice alarms themselves have limitations that must be considered. First, they are likely to be more confusable with (and less discriminable from) a background of other voice communications, such as bell alarm sounds, the task-related communications of dealing with the emergency, or concurrent voice alarms. Second, unless care is taken, they may be more susceptible to frequency-specific masking noise. Third, care must be taken if the meaning of such alarms is to be interpreted by listeners who are less familiar with the language of the voice in a bilingual or even multilingual environment. There is an example of the limitation of voice alarm in 1977. A tragic event occurred at the Tenerife airport in the Canary Islands: A KLM Royal Dutch air lines 747 jumbo jet, accelerating for takeoff, crashed into a Parn American 747 taxiing on the same runway. Although poor visibility was partially responsible for the disaster, the major responsibility lay with the confusion between the KLM pilot and air traffic control regarding whether clearance had been granted for takeoff. An alternative method to solve the problems might be using a redundant system that combines the alerting, distinctive features of the (nonspeech) alarm sound with the more informative features of synthetic voice. Redundancy gain is a fundamental principle of human performance that can be usefully employed [25].

Voice alarm systems are also used in fire alarm and evacuation instruction systems. One of the first evaluations of the design criteria for a voice alarm system was established by McCormick as early as 1964. The design criteria consisted of "audible", "quick-acting", "alerting", "discriminable", "informative", "nonmasking", "nondistracting", and

"nondamaging" [26]. One of the early listings of the characteristics of a voice alarm system was developed at the International Conference on Fire Safety in High-Rise Buildings in 1971. These characteristics consisted of "can give precise instruction under varying emergency conditions", "vary for different zones of the building", can capture attention of the people and alert them to emergencies at hand by preconditioning", "prerecorded voice announcements can be used automatically to respond to manual or automatic fire alarms", etc. [6]. One of the first "vocal alarm system" was designed by Keating and Luftus in the Seattle Federal Building in 1974. The alerting tone consisted of a 1000 Hz, pure sine wave tone, the voice qualities of the communicator were varied with a female voice to initiate the announcement and a male voice to provide the emergency information. The wording of the messages provided three essential contents: (1) Tell the occupants exactly what has happened; (2) tell the occupants what they are to do; (3) tell the occupants why they should do it [26].

Researches on voice alarm signal/systems design have been taken. Liang Zhang *et al.* investigated the effect of speech rate and tune on intelligibility of fire information words and sentences under the conditions with different levels of noise. It is found that the types of signals and noise levels affect the intelligibility significantly, e.g. the appropriate tune for fire information display interface is mezzo-soprano, the appropriate voice rate is 5 characters per second for words display, 7 characters per second for usual sentences and 6 characters per second for the sentences with numbers [12]. Tong Zhang *et al.* studied synthesized speech applied to warn operators of danger or failure in special working environment. The study used speech intelligi-bility test and subjective rating method, and the results showed that the speech rate for voice warnings should be with in the limit of 0.20 second/word and 0.30 second/word, and the optimal speech rate w as 0.25 second/word [13].

The researches above, although conclude in quantitive results, have limitations in giving advice to the design of alarm system and speech signal design for fire or evacuation alarm and instruction system. To improve voice alarm systems design, researches specifically on them are needed.

4 Studies and Observations on Occupants' Response to Voice Alarm

There have been both experimental and observational studies involving practice evacuations and fire incidents. These studies have involved both systems using simple signals and voice notification systems over the past thirty years.

Keating and Loftus [6] conducted two practice evacuations involving 205 participants in Seattle Federal Building in Seattle. The report indicated some general observation of the relocating occupants in response to the voice announcements: 'Personnel unhesitatingly went to the stairwells to evacuate, no one attempted to use the captured elevators, nor was there any pushing, running, or other panicky behavior observed.' Lathrop [7] has reported on a voice alarm system involved in a fire incident in the South Tower of World Trade Center, New York. It should be noted that the validity of any announcement will be questioned by occupants if the announcement is in conflict with direct physical awareness cues. Pigott has described an experiment study in a basement conference room in an office building with two randomly selected adults

groups. The results showed that participants' response to alarm bell are significantly influenced by the voice announcement [8]. Kimura M. Sime J. D. also revealed the effect of authoritative voice instruction through observing college students evacuating from a theatre [9]. Participants who received voice instruction would start to evacuate sooner and take the correct exit. Proulx and Sime conducted 5 evacuations in the Monument underground transit station. Five different kinds of alarm and instruction information (bell only; staff only; public address system telling what to do (PA); staff and PA telling what to do; PA telling what happened, what to do and why) were used respectively in the 5 evacuations. Results showed that the more information the participants got and the more definite the information is, the less the evacuation time is. It also indicated that voice alarm system are most effective in providing comprehensive and intelligible information [27]. Proulx studied practice evacuations in four apartment buildings, one evacuation in each building which has alarm/notification systems consisted of bells. The results indicated the time to initiate the evacuation varied from 30 seconds to 24 minutes, with a mean time of 2.5 minutes [28]. Proulx studied the fire in a high-rise apartment building in Ottawa, Canada, 1997 [29]. It is found that the occupants placed considerable trust in the information received over the voice communication system, for the 83 percent of the respondents who attempted evacuation as a result of evacuation instructions given over this system.

5 Summary and Future Directions

Speech signals and voice alarm systems are more and more implemented in human-computer/machine interfaces and alarm systems. On one hand, rapid advances in microcomputer technology have produced highly efficient speech-synthesis units generating in the recent years. On the other hand, voice alarm systems play an important role in fire alarm and since the first "vocal alarm system" was designed in the Seattle Federal Building in 1974. Previous work has shown that the voice alarm system is effective in facilitating the evacuation by providing information about what happens and what to do [4-9, 11, 27-29]. Studies on both voice alarm and human behavior under emergency conditions with the alarm have been taken. However, there are some limitations that must be considered.

1. In most of the practice evacuations or fire incidents, alarm bell and voice communication are used concurrently. So we can not tell effects of the different signals or communications respectively. Although most of the studies found that voice alarm was better than alarm bell, the effect of different kinds of voice notifications were not clearly described.

2. The previous studies did not investigate differences of human response to voice alarm with different physical references, such as tune, rate, type of voice, content, length of sentences, etc.. So the results of the researches can not provide good reference to the design of voice alarm system.

3. In practice evacuation, noise, smoke and other environmental features under real emergencies are not taken into account, so that the external validity of the results and conclusions is poor.

4. There is a lack of variable (independent variable) control in observations and interview or questionnaire. Experimental method consists of deliberately producing a change in one or more causal or independent variables and measuring the effect of that change on one or more dependent variables. More experimental studies are needed in order to find causal relationships between the voice alarm and occupants' behavior.

5. There is no sound explanation of why voice notification and instruction are effective.

From the discussion above, we can see some future directions in the research areas of human behavior in evacuation, occupants' response to voice alarm systems and ergonomics of voice alarm system design. First, to provide accurate referenced data for improving an emergency action plan or a simulation system, we need quantitive data of human behavior and response to the voice alarm signals in emergency evacuation. Experimental method should be implemented. Second, scientific research is also needed to compare the effectiveness of the human behavioral response of various populations to audible emergency evacuation signal as compared with voice alarm systems and other existing audible signaling systems. Third, in experiments or practice evacuations, effects of physical parameters of voice alarm on human behavior (e.g. response to the alarm, evacuation time) might be investigated. Fourth, features in real emergencies (esp. noise) should be simulated in experiment or practice evacuation. Fifth, multiple methods, including questionnaire, interview, observation, videotape analyzing and experiment, might be combined. Experiment is good at control. Observation and videotape analyzing deals with first-hand resources. Deep interview can help to find reasons of human behavior and more about psychological aspects.

In sum, research on voice alarm system in emergency evacuation is a mutidiscipline area which began ever since 1970s. Various methods have been implemented and valuable results have been concluded. In the future, combination of multiple methods, and usage of new technologies will allow more quantitive, accurate and meaningful results which can provide references to voice alarm system design and give recommendation to evacuation plan and management.

References

1. John, L., Bryan, A.: Selected Historical Review of Human Behavior in Fire. Fire Protection Engineering, Fall, 16 (2002)
2. Fahy, R.F., Proulx, G.: Human Behavior in the World Trade Center Evacuation. International Association for Fire Safety Science. Fire Safety Science. In: Hasemi, Y. (ed.) Proceedings of the Fifth International Symposium, March 3-7, 1997, Melbourne, Australia, Intl, pp. 713–724. Assoc. for Fire Safety Science, Boston, MA (1997)
3. Proulx, G., Sime, J.D.: To Prevent Panic in an Underground Emergency: Why Not Tell People the Truth? In: Fire Safety Science-Proceedings of The Third International Symposium, pp. 843–852. Elsevier, London (1991)
4. Proulx, G., Laroche, C., Latour, J.C.: Audibility problems with fire alarms in apartment buildings. Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting, vol. 2, pp. 989–993. San Diego, Ca (October 9-13, 1995)

5. Proulx, G.: Evacuation Time and Movement in Apartment Buildings. Fire. Safety Journal 24, 229–246 (1995)
6. Keating, J.P., Loftus, E.F.: People Care in Fire Emergencies- Psychological Aspects. Boston: Society of Fire Protection Engineers, Technical Report, 75-4 (1975)
7. Lathrop, J.K.: Two Fires Demonstrate Evacuation Problems in High-Rise Buildings. Fire. Journal 70, 65–70 (1976)
8. Pigott, B.B.: An Administrator's View of Human Behavioural Research from 1975 to 1995. In: Human Behaviour in Fire: Proceedings of The First International Symposium. Fire Sert Centre, University of Ulster, pp. 31–38 (1998)
9. Kimura, M., Sime, J.D.: Exit Choice Behavior During the Evacuation of Two Lecture Theatres. In: Fire Safety Science-Proceedings of The Second International Symposium, pp. 541–550. Hemisphere Publishing Corp, Washington (1989)
10. Mawson, A.R.: Understanding Mass Panic and Other Collective Responses to Threat and Disaster. Psychiatry 68(2), 95–113 (2005)
11. Proulx, G.: Response to Fire Alarms. Fire Protection Engineering http://www.fpemag.com/articles/article.asp?i=267
12. Zhang, L., Sun, X., Zhang, K.: A Research of Speech Signal on Fire Information Display Interface. China Safety Science Journal 16(4), 13–18 (2006)
13. Zhang, T., Zheng, X., Zhu, Z., Jin, W.: A Study on Speech Rate of Voice Warnings. Chinese Journal of Applied Psychology 3(1), 34–39 (1997)
14. Yan, W., Chen, B., Zhong, M.: Research for Different Floors Impacting on College Students' Dispersing Psychology and Actions in the Situation of Fire Disaster. ournal of Safety Science and Technology (Chinese) 1(4), 32–37 (2005)
15. DeCicco, P.R.: Evacuation from Fires, pp. 49–50. Baywood Publishing Company, Amityvile, New York (2002)
16. Tong, D., Canter, D.: The Decision to Evacuate: A Study of the Motivations which Contribute to Evacuations in the Event of a Fire. Fire. Safety Journal 9, 257–265 (1985)
17. Tong, D., Canter, D.: Informative Warnings: In Situ Evacuation of Fire Alarms. Fire. Safety Journal 9, 267–279 (1985)
18. Bellamy, L.J.: Experimental Programme to Investigate Information Warning Characteristics for Motivating Fast Evacuations, Building Research Establishment, Borehamwood, UK (1990)
19. Bellamy, L.J., Geyer, T.A.W., Max-Lino, R., Harrison, P.I., Bahrami, Z., Modha, B.: An Evaluation of the Effectiveness of the Components of informative Fire Warning Systems, Safety in The Built Environment. In: Sime, J. (ed.) SPON, pp. 36–47 (1988)
20. Pigott, B.B.: Fire Detection and Human Behaviour. In: Wakamatsu, T., Hasemi, Y., Seizawa, A., Seeger, P., Pagni, P., Grant, C. (eds.) Proceedings of the Second International Symposium on Fire Safety Science, pp. 573–581 (1989)
21. Klein, R.A.: The Dusseldorf Airport Fire. Fire. Engineers Journal 18–23 (1996)
22. Proulx, G.: Evacuation Times and Movement Times in Apartment Buildings. Fire. Safety Journal 24, 229–246 (1995)
23. Reisser-Weston, E.: Simulating Human Behaviour in Emergency Situations. In: RINA, International Conference of Escape, Fire and Rescue (November 19–20, 1996)
24. Wickens, C.D.: Engineering Psychology and Human Performance, 2nd edn. Harper Collins Publishers, New York (1992)
25. Wickens, C.D., Lee, J., et al.: An Introduction to Human Factors Engineering, pp. 100–102. Prentice Hall, London (2004)
26. Bryan, J.L.: Psychological Variables that May Affect Fire Alarm Design. Fire Protection Engineering 11(Summer 2001)

27. Proulx, G., Sime, J.D.: To Prevent Panic in an Underground Emergency: Why Not Tell People the Truth? In: Fire Safety Science-Proceedings of The Third International Symposium, pp. 843–852. Elsevier, London (1991)
28. Proulx, G.: The Time Delay to Start Evacuating Upon Hearing a fire Alarm. In: Proceedings of The Human Factors and Ergonomics Society. Nshville, pp. 811–815 (1994)
29. Proulx, G.: The Impact of Voice Communication Messages during a Residential High-Rise Fire. In: Human Behaviour in Fire: Proceedings of The First International Symposium. Fire Sert Centre, University of Ulster, pp. 265–274 (1998)
30. Canter, D., Breaux, J., Sime, J.: Domestic, Multiple Occupancy, and Hospital Fires. In: Canter, D. (ed.) Fires and Human Behaviour, 2nd edn. John Wiley & Sons, New York (1990)

Operating Multiple Semi-autonomous UGVs: Navigation, Strategies, and Instantaneous Performance*

Patrik Lif, Johan Hedström, and Peter Svenmarck

Swedish Defence Research Agency (FOI),
Box 1165, SE-58111 Linköping, Sweden
{patrik.lif, johan.hedstrom, peter.svenmarck}@foi.se

Abstract. There is an interest in using multiple unmanned ground vehicles (UGV). The Swedish Army Combat School has evaluated an UGV called SNOOKEN II in a number of field studies. To investigate the possibility to handle multiple vehicles in a simulated setting was set up where the operator simultaneous managed one, two, or three UGVs with limited autonomy. The task was to navigate the UGVs to designated inspection points as fast as possible. The results showed that more inspections were made with multiple UGVs ($p<0.05$), but also that that there was no difference between using two or three UGVs ($p>0.05$). Analysis of use of autonomous mode, route selection, and interviews also show that the subject managed to operate two vehicles with increased performance but that a third vehicle does not provide any extra benefits.

Keywords: UGV, CSE, autonomy, manned-unmanned teaming, multi-robot systems.

1 Introduction

There is currently an increased use of unmanned robotic systems in military operations since they may reduce costs or risk for military personnel, or have enhanced operating characteristics. Unmanned robotic systems are typically used for strategic intelligence, surveillance, and recognizance (ISR). There is, however, a recent interest in also using unmanned systems for tactical situations where manned and unmanned systems operate together as a team [1], [2]. Unmanned robotic systems may for example provide critical information while the manned systems remain in cover or outside of lethal range. One example of such tactical application is the use of unmanned aerial vehicles (UAVs) together with attack helicopters to scout ahead for targets [3], which resulted in an increase in effective weapons range from 5 to over 50 km [4]. Another example is using unmanned ground vehicles (UGVs) for urban combat to peek around corners or surveying areas where the opponent may approach [5].

Since the advantages of UGVs are also of interest to the Swedish Armed Forces, the Swedish Army Combat School (MSS) has for a few years evaluated the UGV

* Corresponding author.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 731–740, 2007.

system SNOKEN II from Saab Aerotech. SNOKEN II has been evaluated in different field studies and proven to be useful in a number of situation, such as gathering of reconnaissance information while minimizing the casualties of own forces and avoiding detection [5]. There is, however, still considerable room for improvement of UGV functionality. Particularly, regarding the operator to vehicle ratio where ideally one operator should be able to control several UGVs without being overloaded [6].

The optimum operator to vehicle ratio depends on several factors, such as task and coordination demands, vehicle autonomy, interface of control station, and the operator's perceptual abilities, working memory, responsibility, and decision making [7]. While some information may be available about a few of these factors, considerable analysis and evaluations are required for any definite recommendations about the operator to vehicle ratio. Such analysis must performed both top-down by considering the overall task and control demands, and bottom-up based on available modes of autonomy and interface capabilities [2]. Generally, higher levels of control become more important as the number of vehicles to operators increase.

A growing body of studies shows that it is a challenging problem to improve the operator to vehicle ratio for UGVs. Since typically only an electro-optical video stream is used for controlling the UGV, the operator faces a remote perception problem in the form of scale ambiguity which makes it difficult to estimate the size of objects, lack of vestibular cues for rate of motion, misperception of rate of motion when the robot's cameras are not at eye height, lack of depth cues, disruption of orientating perceptual functions for where to look next due the reduced field of view, and impoverished control of direction of gaze [8]. Such remote perception problems make it difficult for operators to track their spatial location and identity objects [9]. Controlling a robot using a video stream is not the same thing as being on the scene. In difficult environments, it may not even be possible for one operator to simultaneously control the vehicle and the sensor suite. For example, Burke et al. [10] discuss how fire-fighters that operate a search and rescue robot specialize on either navigation or interpretation of sensor information. Rehfeldt et al. [11] report similar results where two operators perform considerably better than one operator in path planning and target recognition with one or two UGVs. The additional UGV did not improve performance. Further, other attentionally demanding tasks are also hampered when controlling an UGV. Chen and Joyner [12] report how a gunner's performance in detecting and engaging targets decreases significantly when simultaneously controlling an UGV.

Two reasons for why the operator-to-vehicle ratio is so difficult to improve for UGVs are (1) the amount of time a vehicle can be left unattended while maintaining a sufficient performance and (2) the cost for regaining situation awareness when switching between UGVs or other tasks. Crandall et al. [13] show how increasing levels of autonomy for navigation can drastically improve the time between operator interventions. The higher levels of autonomy for navigation in the form of path planning allow better utilization of UGVs for reaching inspection points [14]. Introducing additional UGVs without any autonomy available may in fact decrease the target detection performance in an urban search and rescue task [15]. Similarly, the operator's monitoring performance may also decrease [6], or remain unaffected when more UGVs are used [14].

While some autonomous functions may provide control of a UGV group, UGVs are typically controlled interleaved and sequentially [16]. Since the operator must regain the necessary situation awareness in this mode of control, there is a cost for switching between UGVs. Goodrich et al. [17] show that this cost in the form of time for change detection in visual scene can range from five to twelve second depending on the type of secondary task attended to before returning to the visual scene.

In order to corroborate these results and explore to what extent the operator to vehicle ratio can be improved in a navigation task with only limited UGV autonomy, an experiment was setup in a simulated environment for Military Operation in Urban Terrain (MOUT) where the use of UGVs may be especially advantageous. The operators controlled one, two, or three SNOKEN II, with the only autonomous function of continuing with current speed and heading until stopped by an obstacle. The purpose was to evaluate the control of multiple UGVs from a human-machine interaction perspective, as well as from an overall cognitive system engineering (CSE) perspective [18]. The dependent measures therefore include both objective performance, such as number of inspection points arrived at and route selection, and subjective measures using a simulation interview [19].

2 Method

2.1 Participants

Twelve paid subjects, nine male and three female, with a mean age of 40 years ranging from 25 to 64, participated in the study. All subjects had normal or corrected to normal vision.

2.2 Apparatus

A Dell workstation PWS650 with Intel® Xeon CPU 3.06GHz and 2.0 GB Ram with a Tobii 1750 Eye Tracker integrated into a 17" TFT monitor was used to run the simulation in the MSI laboratory at FOI Linköping, Sweden. A Saitek Cyborg Evo joystick was used to manually control the UGV, control the camera, engage the autonomous mode, and switch control between UGVs.

2.3 Design and Stimuli

In a simulated MOUT environment, the participants operated one, two and three UGVs simultaneously, to evaluate their ability to control multiple UGVs. As in the study by Trouvain et al. [14], the task was to drive the UGVs as quickly as possible to designated inspection points in the form of color coded inspection points, blue, red, and green, for each UGV respectively. Figure 1 shows the interface of controlling the UGVs. The inspection points were presented both in the primary 3D environment and at the 2D map in upper right corner. The map could be zoomed out for an overview (default view) or zoomed in to reveal more details about the terrain. Once a UGV arrived at the inspection point, the next inspection point was presented. The UGVs could be controlled manually with a joystick, set in an autonomous mode to continue with the current speed (maximum 14 km/h) and heading, or a camera mode where the

UGV was standing still and the operator only rotated a camera to obtain information of the surroundings. The autonomous mode was activated by pressing a button on the joystick and deactivated by pushing the same button or by manipulating the joystick in any direction and another button to switch between UGVs. Subjects were completely free to choose how to control the UGVs. The design included three 10 minutes blocks (one, two, and three UGVs) which were counterbalanced between subjects.

Fig. 1. Interface used in the simulated 3D virtual environment and 2D overview map presentation

The research questions regarding the use of multiple UGVs were;

- How is the performance affected, such as the number of inspections and the percentage of time when the UGVs are at a standstill?
- How is the use of the autonomous mode affected?
- How are the strategies affected?

Both objective and subjective measures were collected for an overall understanding of how the subjects controlled multiple UGVs. The objective measures were the number of inspections, time spent in different control modes, and instantaneous performance (IP). A complementary subjective measure was also collected with a simulation interview [19]. The motivations for these dependant measures are:

Number of inspections): to evaluate if multiple UGVs help the operator reach more inspection points. If the operator manages to control multiple UGVs by simultaneously using manual and autonomous modes, more inspections will be performed with two, and still more inspections with three UGVs. If the task is too demanding and cause high mental workload, the operator will not perform more inspections with multiple UGVs.

Time spent in different control modes: to evaluate the subjects control strategy. The control modes are either attended when an UGV is controlled in the primary 3D window or unattended when another UGV is controlled in the primary window. For both attended and unattended control modes there were four possible modes, manual,

autonomous, autonomous stop, and camera. This means that each UGV can be in one of eight possible modes as shown in Table 1. It is important to understand that an UGV also can be efficiently used when unattended in autonomous mode. In order to improve performance it is necessary to use multiple UGVs, and to assure that the UGVs that are not controlled in the primary window are advancing towards the inspection points in autonomous mode.

Table 1. Each UGV has eight possible modes

	Attended (controlled in primary window)	**Unattended (controlling another UGV)**
Manual	Manually controlled	At a standstill
Autonomous	Autonomously maintains speed and heading while controlled in the primary window	Autonomously maintains speed and heading while another UGV is controlled in the primary window
Autonomous stop	At a standstill after visibly stopping for an obstacle in autonomous mode	At a standstill from an unintentional stop against an obstacle
Camera	Active use of the camera	At a standstill in camera mode

Instantaneous Performance (IP): to evaluate the efficiency of the subjects control strategy. Crandall et al. [13] propose IP as a measure for the extent to which a robot travels the shortest path towards the target (inspection point) at maximum speed. The A* algorithm was used to calculate the shortest path in this experiment. IP is calculated by dividing the distance actually covered relative the shortest path (instantaneous work) with the distance that could have been covered at maximum speed (instantaneous capacity). This means that IP is always a value between -1 and 1, where 1 means that the robot travels optimally towards the target at the highest possible speed and -1 optimally from the target. If IP is zero, the robot is at a standstill. IP is used to evaluate each subject's performance with one and multiple UGVs. If the subjects can handle multiple UGVs they will maintain a good IP for all UGVs. On the other hand, if the performance is high for one UGV and low for the others, this indicates an interleaved sequential control. The IP was calculated in one second intervals.

Simulation interview: to get an understanding of the operators view on solving the task at hand, as well as detailed information about their cognitive processes. The subjects first describe the major tasks in the situation and develop each event to describe actions, situation assessments, critical cues, and potential errors that an inexperienced person would be likely to make [19]. These interviews complement the objective measures and widen the understanding of using multiple UGVs in urban terrain.

2.4 Procedure

The participants where informed about the experiments purpose, and allowed to ask clarifying questions. They were then instructed on how to use the buttons on the

joystick to control the UGV and camera, and also how to switch between UGVs. All participants were then trained for 20 minutes, using one single UGV and multiple UGVs. After training, the experiment was conducted with three counterbalanced sets (number of UGVs). Each set was stopped after 10 minutes whether all inspection points were reached or not. Finally, a simulation interview was performed, where the participants explained events, actions, critical cues, control strategies, and potential errors.

3 Results

A repeated measures ANOVA shows that the subjects performed more inspections with multiple UGVs ($F(2, 22) = 5.27$, $p < 0.05$). There was no significant difference between using two or three UGVs ($p > 0.05$). Figure 2 shows that with one UGV, a mean of about six inspections were performed, and with two or three UGVs about 8 inspections were performed.

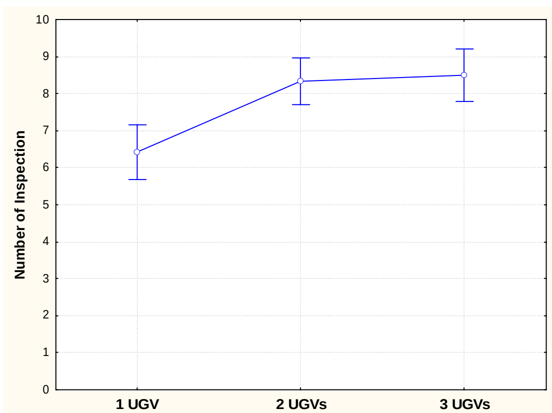


Fig. 2. Mean number of inspections when using one, two, and three UGVs simultaneously. The bars denote ± Standard Errors.

The time spent in the attended control modes was analyzed by calculating the percentage of time in each of the control modes since the total time for attended control was the same irrespectively of the number of UGVs. The time spent in unattended control modes, on the other hand, was analyzed by calculating the average percentage of time in each control mode for all UGVs. Only control modes that were used reasonably frequent were analyzed further. These control modes were attended manual and autonomous, unattended manual, autonomous, and autonomous stop. The camera mode was seldom used since it was not really required for this task. One subject was excluded from the analysis of time in control modes due to problems with the data collection.

There was a significant effect of the number of UGVs on the percentage of time using manual attended control ($F(2, 20) = 4.43$, $p < 0.05$). However, the differences

are small and on average the UGVs are controlled manually about 90% of the time. This leaves only 10% of the time for control in the other attended control modes, such as attended autonomous.

The percentage of time using the autonomous unattended mode increase dramatically from one to multiple UGVs ($F(2, 20) = 33.0$, $p < 0.0001$). There was, however, no difference between using two and three UGVs. Figure 3 shows that when the autonomous unattended mode is used about 20% of the time when controlling two or three UGVs. Only about 7% of the time was, however, used for supervising the UGVs in autonomous attended mode irrespectively of the numbers of UGVs ($p > 0.05$). The UGVs can, on the other hand, also be partly supervised on the map where the operator can see if an UGV in moving in the correct direction or if it is standing still. This control strategy was, however, not very efficient since the percentage of time with unintentional stops, that is unattended autonomous stop, increase dramatically when using multiple UGVs, ($F(2, 20) = 18.3$, $p < 0.0001$). Figure 3 shows that the percentage of time with unintentional stops is about 13% with two UGVs and increase to about 18% with three UGVs.

The percentage of time where the UGVs stand still without any direction, that is manual unattended mode, increase when using multiple UGVs ($F(2, 20) = 26.5$, $p < 0.0001$). Figure 3 shows how the manual unattended mode increases to about 17% with two UGVs and to about 28% with three UGVs. The analysis of unintentional manual and autonomous stops shows that the mental workload is high.

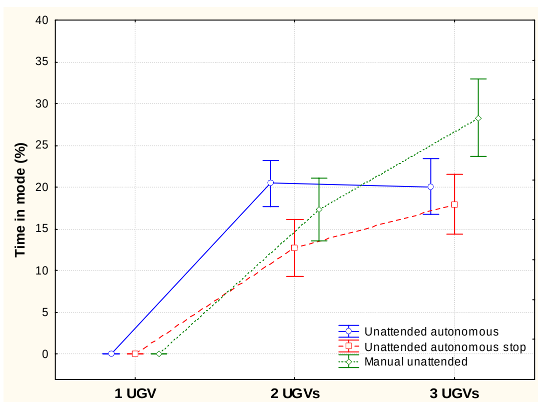


Fig. 3. Percentage of time when UGVs are unattended in manual, autonomous, or autonomous stop mode. The bars denote the ± Standard Errors.

Although the number of inspections show the overall performance it is also interesting to investigate alternative performance measures that provide more information about the subjects control strategies. The IP provides such a measure by combining the distance traveled relative the shortest path and the velocity relative the highest possible speed. The IP was calculated for each UGV once per second. For the statistical analysis, the average IP was calculated for the whole scenario and all UGVs. A repeated measures ANOVA shows that the average IP decreases with multiple UGVs ($F(2, 20) = 24.1$, $p < 0.0001$). There was no significant difference

between using two or three UGVs ($p > 0.05$). A correlation analysis between the average IP and the average number of inspections per UGV show a significant relationship ($r = 0.95$, $p < 0.05$). The analysis shows that IP potentially can explain subjects control strategies in way that predict the number of inspections.

An exploratory investigation of IP clearly reveals a pattern where the operator control one UGV with acceptable performance (mean 0.46) which then decline when two and especially three UGVs are controlled simultaneously (mean 0.27 and 0.17 for each UGV). Figure 3 shows an example of one subject's IP when controlling three vehicles simultaneously. Lowess smoothing method was used to fit a curve to the data. A performance of 1 indicates that the UGV has a perfect direction with maximum speed towards the inspection point, 0 that the UGV is not moving, and -1 that the UGV has maximum speed in the opposite direction. Figure 3 shows that the subject only manages to have moderate or high performance for one UGV at a time. The UGVs performance quickly increases with attentional control, but decreases rapidly when the attention shifts to another UGV.

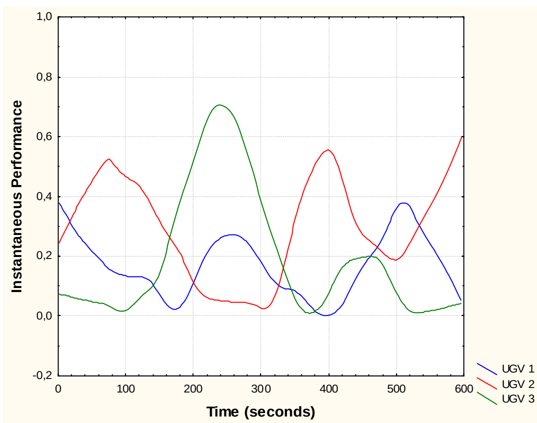


Fig. 4. Instantaneous performance for one subject controlling three UGVs simultaneously. The figure shows that an acceptable IP can only be maintained for one UGV at a time.

The simulation interviews show that the subjects generally divide the overall task into three control tasks, (1) manual control and transition to autonomous mode, (2) navigation, and (3) utilization of the UGVs. Manual control consists of maintaining desired speed and heading while considering the local terrain in the form of obstacles, passability, and inclination. The subjects typically consider both the local terrain and map before activating the autonomous mode when controlling multiple UGVs. The subjects only activate the autonomous mode on straight roads without obstacles. Several subjects use a strategy of directing the UGV towards a wall, when placed in autonomous mode. This way, the UGV will at least stop at a specific location even if they are not able to supervise it again before it reaches the wall. This can be seen as a safety maneuver and an understanding of that the mental workload is high. The subjects try to supervise the unattended UGVs on the map in order regain control when needed. The map is normally zoomed out for an overall understanding, but may

be zoomed in for more detailed information in specific situations, such as additional information about passability.

Concerning navigation, the subjects mainly prefer main roads and avoid small tricky passages. This increases the opportunity to use the autonomous mode. The subjects also prioritize the inspection points for multiple UGVs depending on their distance and accessibility.

Finally, the subjects' main strategy to utilize multiple UGVs was to use all vehicles simultaneously. In some situations, however, most subjects had to abandon one or two UGVs. Some subjects prefer to manually control one UGV, supervise another in autonomous mode, and abandon the third UGV when the mental workload is high.

4 Discussion and Conclusions

The results show that subjects can efficiently control one UGV and that the performance increases with two UGVs. A third UGV does not provide any additional benefits, however, since the subjects are already overloaded. Due to the high mental workload, unattended UGVs are often at a standstill awaiting commands, sometimes just for a few seconds but sometimes up to one minute. This is a problem from a tactical and safety point of view since an unattended UGV could be tampered with and to fully use the UGV potential it must be controlled by the operator at all times. A higher performance in this navigation task can only be achieved with more sophisticated autonomous functions or a control interface that allows more efficient supervision and reduces the switching costs when alternating the attended control between UGVs. For example, Trouvain et al. [14] show how autonomous path planning functions can improve the utilization of multiple UGVs. Unfortunately, more sophisticated automation may increase the task switching costs, which can be mitigated if operator can choose the level of control [20]. Future studies will investigate these issues for advancing in urban terrain with multiple UGVs and support from autonomous functions, such as area search for mobile threats.

References

1. Barnes, M.J., Cosenzo, K.A., Mitchell, D.K., Chen, J.Y.C.: Human Robot Teams as Soldier Augmentation in Future Battlefields: An Overview. In: Proceedings of HCI International - 11th International Conference on Human Computer Interaction (2005)
2. Svenmarck, P.: Principles of Human-Robot Coordination for Improved Manned-Unmanned Teaming. FOI Memo 1507. FOI – Swedish Defence Research Agency, Linköping, Sweden (2005)
3. Fayaud, G.R.: The airborne manned unmanned system. Unmanned Systems 19(4), 16–21 (2001)
4. Army pushing helo-TUAV integration. Rotor & Wing (May 2001)
5. Lif, P.J., Jander, H., Borgvall, J.: Tactical Evaluation of an Unmanned Ground Vehicle During a MOUT Exercise. In: Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, San Francisco, CA (2006)

6. Chadwick, R.A.: Operating Multiple Semi-Autonomous Robots: Monitoring, Responding, Detecting. In: Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, San Francisco, CA (2006)
7. Uninhabited Military Vehicles (UMVs): Human Factors Issues in Augmenting the Force. RTO-TR-HFM-078. NATO/RTA (2006)
8. Tittle, J.S., Roesler, A., Woods, D.D.: The Remote Perception Problem. In: Proceedings of the Human Factors and Ergonomics Society 46th annual meeting, Baltimore, MD, pp. 260–264 (2002)
9. Darken, R.P., Kempster, K., Peterson, B.: Effects of Streaming Video Quality of Service on Spatial Comprehension in a Reconnaissance Task. In: Proceedings of the Interservice/Industry Training Simulation Education Conference, Orlando, FL (2001)
10. Burke, J.L., Murphy, R.R., Coovert, M.D., Riddle, D.L.: Moonlight in Miami: An Ethnographic Study of Human-Robot Interaction in the Context of an Urban Search and Rescue Disaster Response Training Exercise. Human-Computer Interaction 19, 85–116 (2004)
11. Rehfeld, S., Jentsch, F., Curtis, M., Fincannon, T.: Proceedings of HCI - 11th International Conference on Human Computer Interaction, Las Vegas, NV (2005)
12. Chen, J.Y.C., Joyner, C.T.: Individual Differences in Concurrent Performance of Gunner and Robotic Operator’s Task in a Multitasking Environment. In: Human Factors and Ergonomics Society 50th Annual Meeting. San Francisco, CA, pp. 1759–1763 (2006)
13. Crandall, J.W., Goodrich, M.A., Olsen, D.R., Nielsen, C.W.: Validating Human-Robot Interaction Schemes in Multi-Tasking Environments. IEEE Transctions on Systems Man, and Cybernetics, Part. A (4), 438–445 (2005)
14. Trouvain, B., Wolf, H.L., Schneider, F.E.: Impact of Autonomy in Multirobot Systems on Teleoperation Performance. In: Schultz, A.C., Parker, L.E., Schneider, F.E (eds.) Multi-Robot Systems: From Swarms to Intelligent Automata. Proceedings from the 2003 International Workshop on Multi-Robot Systems, vol. II, Kluwer, Dordrecht (2003)
15. Chadwick, R.: Multiple Robots and Display Views: An Urban Search and Rescue Simulation. In: Proceedings of the Human Factors and Ergonomics Society, Orlando, FL, pp. 387–391 (2005)
16. Chadwick, R.A., Gillan, D.J., Simon, D., Pazuchanics, S.: Cognitive Analysis Methods for Control of Multiple Robots: Robotics on $5 a Day. In: Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting, pp. 688–692. Santa Monica, CA, (2004)
17. Goodrich, M.A., Quigley, M., Cosenzo, K.: Task Switching and Multi-Robot Teams. In: Parker, L.E., Schneider, F.E., Schultz, A.C. (eds.) Multi-Robot Systems. From Swarms to Intelligent Automata. Proceedings from the 2005 International Workshop on Multi-Robot Systems, vol. III, Springer, Dordrecht (2005)
18. Hollnagel, E., Woods, D.D.: Joint Cognitive Systems. Foundations of Cognitive Systems Engineering. Taylor & Francis, Boca Raton, FL (2005)
19. Militello, L.G., Hutton, R.J.B.: Applied Cognitive Task Analysis (ACTA): A Practitioner’s Toolkit for Understanding Cognitive Task Demands. Ergonomics 41(11), 1618–1641 (1998)
20. Squire, P., Trafton, G., Parasuraman, R.: Human Control of Multiple Unmanned Vehicles: Effects of Interface Type on Execution and Task Switching Times. In: Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, pp. 26–32 (2006)

Evaluation of the Effects of Visual Field on Road Sign Recognition

Bor-Shong Liu[1], Chih-Hung Hsu[2], Hsien-Yu Tseng[1], and Tung-Chung Chia[3]

[1] Department of Industrial Engineering and Management, St. John's University
Sec. 4, Tam King Road, Tamsui, Taipei 25135, Taiwan
bsliu@mail.sju.edu.tw
[2] Department of Industrial Engineering and Management, Hsiuping Institute of Technology, Taichung 41249, Taiwan
[3] Department of Business Administration, Ling Tung University

Abstract. The objective of this study was utilized the head and eye-tracker system to examine the effects of visual field between sedan and mixed concrete truck on road sign recognition. Participants were asked to search the traffic signs on driving video and give a verbal report instantaneously for visual search. Results of ANOVA revealed that the mode of vertical direction of eye gaze were significant differences between two conditions. Lower eye gazes were found in higher visual field from mixed concrete truck than in sedan. In addition, subjects reported fast in video of mixed concrete truck than in sedan. According the fixations data, subjects search quickly on painted speed limit on road in mixed concrete truck as compared to search the post or cantilever speed limit sign in video of sedan. These findings can be used or further application on signage design, vehicle design, and driver support information system design.

Keywords: Driving behavior; Visual search; Truck; Traffic signs.

1 Introduction

Every day, millions of people drive to work, to shop, to conduct business, or to be entertained. Tracking and continuous manual control are normally a critical part of any human-vehicle interaction. Visual search and decision making are the two primary components that are central to visual inspection tasks [9]. Further, visual search can be defined as the process of locating a target/defect within the area of interest and decision making can be defined as the process of deciding whether the target/defect is acceptable or not [19]. Drury [8] has defined visual search as a series of fixations, linked by eye movements, during which information is captured for further processing. Eye movements are necessary to search the visual field [11]. Further, eye movements can generally be divided into two major classes. Pursuit movements are those of constant velocity that are designed to follow moving targets (e.g. following the vehicle across the intersection). More related to visual search are saccadic eye movements, which are abrupt discrete movements from one location to the next. The actual movement time is generally quite fast (typically less than 50 ms)

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 741–749, 2007.

and is not much greater for longer than shorter movements. The destination of a scan is usually driven by top-down process (i.e. expectancy), although on occasion a saccade may be drawn by salient bottom-up process (e.g. flashing light). In addition, the dwell duration is governed jointly by two factors: (1) the information content of the item fixated (e.g. when reading, longer words require longer dwells than short ones), (2) the ease of information of extraction, which is often influenced by stimulus quality [27].

Many studies have shown that attention and eye movements are guided by a common selection process [1]. Hughes and Cole [12, 13] state that the driver's information needs should have a significant effect on visual search strategy. Visual search strategies can be broadly classified into two types: random search and systematic search. Random search is defined as a memoryless process in which each fixation is equally likely to occur anywhere in the search field [2]. Systematic search is defined as a process with perfect memory. Thus, the same location is never inspected more than once [28]. Mori and Abdel-Halim [18] found that under free driving conditions without a specific search instruction only 11.2% of the traffic signs were fixated for a duration that was equally long compared to the condition in which participants were instructed to read all signs. Luoma [15] suggested that drivers look at virtually every traffic sign, with warning signs being fixated twice on average, with average glance duration of 500 ms. Groeger et al. [10] have been research carried out to investigate the visual behavior of train drivers. The study showed that approximately 50% of time approaching signals was spent scanning the visual scene. The remaining time was spent fixating on railway signage and infrastructure, locations beside the track and signals. The Wickens' visual sampling model stated that safety critical targets would receive greater visual sampling than non-safety critical targets [26]. In addition, the experiment was conducted to determine whether the decrease in fixation times for traffic signs, information signs or road markings and found main effect of traffic object on the total fixation time and on fixation frequency. [16]. Crundall and Underwood [6] reported that the case of novices driving on a dual-carriageway showed drivers were searching along the horizontal meridian no differently on roads than on a relatively quiet rural road. Further, Underwood et al. [23] indicated that experienced drivers showed more extensive scanning on the demanding sections of dual-carriageway.

Search models can be extremely important in human factors and eye movement research has been the basis for many applications [5, 19]. Many studies have shown the effects of drivers' characteristics (e.g. trained vs. novice), traffic signs and in-vehicle task on visual search patterns [4, 16, 24]. However, relatively few studies have conducted to explore the effects of vehicle type on visual search patterns. Thus, in the present investigation, head and eye-tracker system were utilized to examine the effects of visual field between sedan and mixed concrete truck on visual search patterns.

2 Method

2.1 Participants

Twelve participants were recruited as paid volunteers for this study. The age of the subjects ranged from 25 to 45 years, average age was 35.2 years. The sample

consisted of six females and six males. All participants had 0.8 corrected visual acuity or better and normal color visual. In addition, all subjects had held a driving license for at least two years and drove 5,000 km or more annually.

2.2 Apparatus

The head and eye-tracker system (faceLab 4.2, Seeing Machines, Australia) was used to monitor the driver's eye gaze and record the duration of fixation on specific objects. This information was used to determine whether the subject took in all the information or only part of it, whether the subject comprehended the information. The Stereo-head with pair of Sony cameras (FCB-EX480B) mounted the infra-red pods. Before experiment, two driving scenes were recorded from the driver’s viewpoint on the same route in both sedan and mixed concrete truck (TRV30, Sony, Japan). Further, the videos were captured and saved as AVI (Audio Video Interleave) format by Movies Maker Program. The videos were played back by RealPlayer 10.5 program and projected on a video screen by Plus XGA beamer with 1204*768 pixels and 1100 lm (Plus, U31100W, Japan). The projection visible screen had a width of 120 cm and height of 84 cm. Participants were seated approximately 100 cm from the projection screen (Fig. 1).

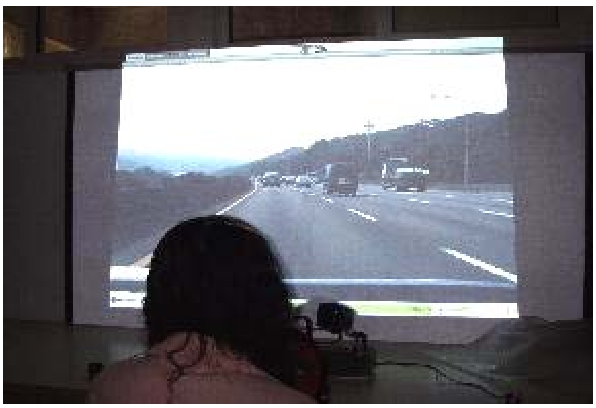


Fig. 1. Driving video projected on a video screen

2.3 Procedure

With tracker system running (faceLab), it is important to correctly position the Stereo-head. This involves two considerations: the vergence of the cameras within Stereo-head, the location the Stereo-head at the bottom edge of the screen, aligned horizontally with the center of the screen and the tilt angle is less than 15 degrees. A Stereo-head must be recalibrated every time while positions change, or the camera zooms are changed. There are three steps to calibrating the cameras including capturing data for calibrating, processing the calibration and installing the cameras parameters into faceLab. Capturing data for camera calibration involves holding the calibration target in various poses in view of the cameras, while faceLab takes a number of image snapshots for analysis [21].

Before the actual experiment was conducted, two driving scenes were recorded on the same route from the viewpoint of a driver in both vehicles (i.e. sedan and mixed concrete truck). Further, the videos were captured and saved as AVI (Audio Video Interleave) format by Movies Maker Program. Upon arrival, the participants were given general instructions about the experiment. Each participant sat behind the steering wheel and had a clear view on the video screen. The driving video sequence was presented to all subjects in random order. While the video started, participants were asked to search the traffic signs (e.g. speed limit) until the video stopped and had verbal report instantaneously about results of visual search (Fig 2). If the drivers experienced difficulties in making verbal reports and only provided brief verbalizations, the experimenter prompted the driver with probe questions such as: “what are you looking at here? Why are you focusing on that particular feature?“ In addition, an experimenter seated in the right recorded the reaction time.

a

b

Fig. 2. Two video conditions from (a) sedan and (b) mixed concrete truck

2.4 Data Analysis

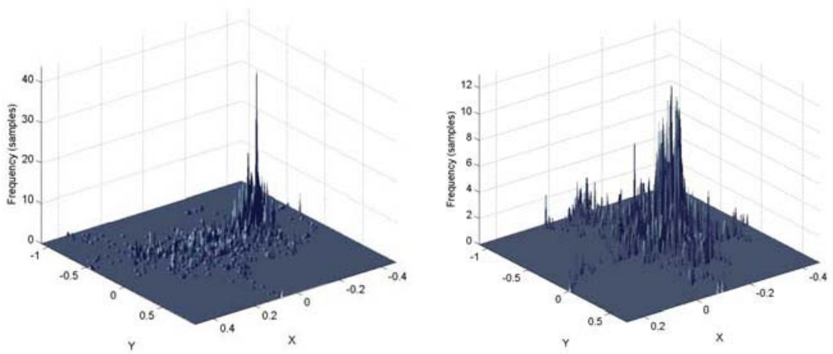
All experimental measures can be saved to a disk logfile with the directory base name (directorybasename)_(date and time) (e.g. Facelab_Fri_Feb_2_2006_10_32_36). There are five different logfiles which include the filename prefixes: Eye, Features, Head, Timing, and World. In the World Model, the faceLab system can store the eye position values corresponding to all the target positions on screen, in relation to an X-Y coordinate axis. Further, data files can be converted from binary mode into ASCII format for analysis by Data Browser 2.1 (faceLab, Seeing Machines). In further analysis, the number of fixations per cell were measured when the screen is divided into several cells according the square (100×100 pixels). Apart from data of eye gaze, performance of visual search have been measured in time (second). Analysis of variance (ANOVA) was utilized to determine the effects of visual field on dependent variables.

3 Results

3.1 Eye Gaze Angles

Fig 3 showed the eye gaze within pitch and yaw angle corresponding X and Y coordinates on screen. For sedan condition, the eye pitch angles (X axis; up-down

rotation) ranged form -0.46 (down) to 0.47 (up) radians and the mode is -0.17 radians. The eye yaw angles (Y axis; left-right rotation) ranged from -0.68 (left) to 0.93 (right) radians and the mode is -0.59 radians. For mixed concrete truck conditions, the eye pitch angles ranged form -0.46 (down) to 0.25 (up) radians and the mode is -0.44 radians. The eye yaw angles ranged form -1.1 (left) to 0.7 (right) radians and the mode is -0.25 radians. Results of eye gaze angles revealed that the vertical scanning areas were larger in sedan condition (0.93 radians) than in mixed concrete truck (0.71 radians). On the other hand, the horizontal were larger in mixed concrete truck (1.8 radians) than in sedan condition (1.61 radians). In addition, the modes of eye gaze angles were further analyzed. Results of ANOVA revealed that the mode of pitch angles of eye gaze (vertical direction) were significant difference between conditions ($p < 0.05$). Modes of eye pitch angles were lower in mixed concrete truck (-0.171 radians) than in sedan (-0.012 radians). For horizontal scanning pattern, results of ANOVA revealed that modes of yaw angles of eye gaze were not significant difference between to two conditions. There are -0.337 and –0.081 radians in mixed concrete truck and sedan conditions, respectively.



Sedan Mixed concrete truck

Fig. 3. Eye gaze angle: Sedan (left), Mixed concrete truck (right)

3.2 Distribution of Fixations

Fig 4 showed the distribution of fixations within corresponding X and Y coordinates on screen. For video from sedan, fixations of subject focused on center of screen (coordinate X=0, Y=0) and vertical pattern of visual search. By contrast, fixations of subjects focused on below area of screen in video from mixed concrete truck and they trend horizontal pattern of visual search. The data can be analyzed by considering an increasing number of squares. In addition, more detailed analysis of observations, with data from squares cells. There are three areas with longer fixation times in video of sedan. One of these corresponds to cell 7, -7 (X=701~799, Y= -701`-799), cell 8, -7 (X=801~899, Y= -701~-799) and cell 4, -7 (X=401~499, Y= -701~-799), with 319, 304 and 174 fixations, respectively. For video of mixed concrete truck, there are cell 9, -6 (X=901~999, Y= -601~-699), 10, -7 (X=1001~1099, Y= -701~-799) and 8, -6 (X=801~899, Y= -601~-699), with 187, 152 and 146 fixations, respectively.

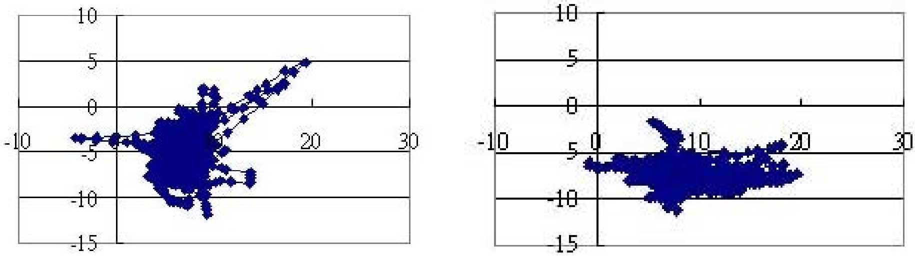


Fig. 4. Scatter plot of gaze direction vector screen intersection X and intersection Y

3.3 Performance of Visual Search

Participants were asked to search the traffic signs until the video stopped and had verbal report instantaneously about results of visual search. For example, what is the speed limit now (Fig 5)? The report time was significant difference between the two video conditions ($p < 0.05$). Subjects reported fast in video of mixed concrete truck (8.53 sec) than in video of sedan (13.26 sec). In addition, according the fixations data, subjects search quickly on painted speed limit on road in video of mixed concrete truck as compared to search the post or cantilever speed limit sign in video of sedan, because the visual field is higher and wider on mixed concrete truck than on sedan.

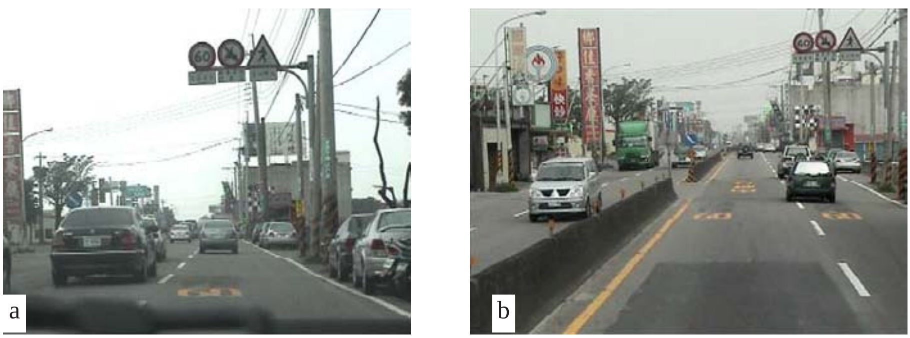


Fig. 5. Speed limit sign (60) and painted on road. (a) viewpoint from sedan and (b) mixed concrete truck.

4 Discussion

Looking in the wrong direction or taking your mind off driving at a critical moment can lead to disastrous consequences while driving [24]. Driver inattention or distraction is estimated to be responsible for 25-30% of police-reported traffic crashes, 1.2 million crashes per year in the United States, according to the National Highway Traffic Safety Administration (NHTSA) [22]. From a road safety perspective it is common sense that a driver must look at the appropriate location in a traffic scene in order to gain information about risks and potential risks in the scene.

Many studies have recorded and analyzed drivers' eye movements. At the visual field becomes more complex, the number of eye movements made increases, and the mean fixation duration decreases [17]. Martens and Fox [16] found that there was a main effect of traffic object on the total fixation time and on fixation frequency. The warning speed camera, start urban areas Soest, and warning pedestrians were higher fixation times.

Luke et al. [14] found that signal types have a strong effect on visual behavior. When train drivers were approaching gantry signals they looked at the signal more often, for longer, and for a greater percentage of the time overall compared to when approaching post or cantilever signals. In addition, Groeger et al. (2003) also reported that drivers appeared to glance at gantry signals earlier than post or cantilever signals.

Hughes and Cole [12] instructed drivers to verbally report all the objects that attracted their attention. Half of the participants watched a 16 mm colour movie file made from the route and half of the participants actually drove a vehicle in this environment. The verbal report data indicated that in comparison to real driving a movie provided a reasonably adequate simulation. There are a number of interactions between the type of road environment and the events occurring. Mean fixation durations tend to be longer when viewing films of rural roads than urban ones [3]. While the first effect may be driven purely by the visual characteristics of the scene, the second presumably reflects the driver's knowledge about danger and the additional time spent processing such information. In present study, subjects asked to watch video-recordings taken from two types vehicles (i.e. sedan and mixed concrete truck) traveling along a range a road. Results of analysis for eye gaze showed that the visual search patterns are significant difference between scenes. Fixations of subject focused on below area of scene from mixed concrete truck and they trend horizontal pattern of visual search. While searching the speed limit on road, for example, subjects search quickly on painted speed limit on road in video of mixed concrete truck as compared to search the post or cantilever speed limit sign in video of sedan, because the visual field is higher and wider on mixed concrete truck than on sedan. In addition, much research has shown that drivers look to a future-path, road center region a few degrees down from true horizon about 80-90% of the time and only rarely at lane markings near the vehicle during normal driving [20, 25].

5 Conclusions

Drivers adapt their eye movement behavior to the driving environment. In general, they increase viewing time in the road center area when driving task difficulty increases, as evidenced by spatial gaze concentration being highest in the rural curves, followed by rural straight sections, and motorway [24]. On the other hand, a visibility issue pertains to signage [7]. Both searching for and reading critical highways signs can be a source of visual distraction. Hence, there is an important need for highway designers to: (1) minimize visual clutter from unnecessary signs, (2) locate signs consistently, (3) identify sign classes distinctly, and (4) allow verbal signs to be read efficiently by giving attention to issues of contrast sensitivity and glare. Present investigation were utilized the head and eye-tracker system to examine the effects of visual field between sedan and mixed concrete truck on visual search patterns. Results

of analysis for eye gaze indicated that various visual filed between two vehicle types affected strong the visual pattern. Differences between various vehicle types should be taken account in planning process. These findings can be used or further application on signage design, vehicle design, and driver support information system design.

Acknowledgments. This study is supported by a grant from the National Science Council, Taiwan, ROC, and Project No. NSC 94-2213-E-129-006**.**

References

1. Bichot, N.P.: Attention, Eye Movements, and Neurons: Linking Physiology and Behaviour. In: Jenkin, M., Harris, L. (eds.) Vision and Attention, Springer, New York (2001)
2. Bloomfield, J.R.: Theoretical Approaches to Visual Search. In: Drury, C., Fox, J. (eds.) Human Reliability in Quality Control. Taylor and Francis, London (1975)
3. Chapman, P., Underwood, G.: Visual Search of Driving Situations: Danger and Experience. Perception 27, 951–964 (1998)
4. Chapman, P., Underwood, G., Roberts, K.: Visual Search Patterns in Trained and Untrained Novice Drivers. Transportation Research Part. F 5, 157–167 (2002)
5. Chi, C.F., Lin, F.T.: A New Method for Describing Search Patterns and Quantifying Visual Load Using Eye Movement Data. International Journal of Industrial Ergonomics 19, 249–257 (1997)
6. Crundall, D., Underwood, G.: Effects of Experience and Processing Demands on Visual Information Acquisition in Driving. Ergonomics 41, 448–458 (1998)
7. Dewar, R.: Warning: Hazardous Road Signs Ahead. Ergonomics in Design 26–31 (1993)
8. Drury, C.G.: Inspection Performance. In: Salvendy, G. (ed.) Handbook of Industrial Engineering, pp. 2282–2314. Wiley, New York (1992)
9. Drury, C.G., Clement, M.R.: The Effect of Area, Density, and Number of Background Characters on Visual Search. Human Factors 20, 597–602 (1978)
10. Groeger, J.A., Bradshaw, M.F., Everatt, J., Merat, N., Field, D.: Pilot study of train drivers' eye-movements. University of Survey Technical Report for Rail Safety and Standards Board, London (2003)
11. Hallett, P.E.: Eye movements. In: Boff, K.R., Kaufman, L., Thomas, J.P. (eds.) Handbook of Perception and Human Performance, vol. 1, Wiley, New York (1986)
12. Hughes, P.K., Cole, B.L.: What Attracts Attention When Driving? Ergonomics 29, 337–391 (1986)
13. Hughes, P.K., Cole, B.L.: The Effect of Attentional Demand on Eye Movement Behaviour When Driving. In: Gale, A.G., et al. (ed.) Vision in Vehicles II, pp. 221–230. Elservier, New York (1988)
14. Luke, T., Brook-Carter, N., Parkes, A.M.: An Investigation of Train Driver Visual Strategies. Cognition, Technology & Work. 8, 15–29 (2006)
15. Luoma, J.: Immediate Responses to Road Signs of Alerted and Unaltered Drivers: An Evaluation of the Validity of Eye Movement Method. In: Presented at the Transportation Research Board annual meeting, Washington, DC (January 1992)
16. Martens, M.H., Fox, M.: Does Road Familiarity Change Eye Fixations? A Comparison Between Watching a Video and Real Driving. Transportation Research Part. F 10, 33–47 (2006)

17. Miura, S.: Active Function of Eye Movement and Useful Field of View in a Realistic Setting. In: Groner, R., et al. (ed.) From Eye to Mind Information Acquisition in Perception Search and Reading, pp. 119–127. Elsevier, Amsterdam (1990)
18. Mori, M., Abdel-Halim, M.H.: Road Sign Recognition and Non Recognition. Accident Analysis and Prevention 13, 101–115 (1981)
19. Nalanagula, D., Greensein, J.S., Gramopadhye, A.K.: Evaluation of the Effect of Feedforward Training Displays of Search Strategy on Visual Search Performance. International Journal of Industrial Ergonomics 36, 289–300 (2006)
20. Recarte, M.A., Nunes, L.M.: Effects of Verbal and Spatial-imagery Tasks on Eye Fixations While Driving. Journal of Experimental Psychology: Applied 6, 31–43 (2006)
21. Seeing machines. FaceLab 4 user Manual. Seeing Machines, Australia (2005)
22. Shelton, L.R.: Statement Before the Subcommittee on Highways and Transit, Committee on Transportation and Infrastructure, U.S. House of Representatives (2001). Available from http://www.nhtsa.dot.gov/nhtsa/announce/testimony/distractiontestimony.html
23. Underwood, G., Chapman, P., Bowden, K., Crundall, D.: Visual Search While Driving: Skill and Awareness During Inspection of the Scene. Transportation Research Part. F 5, 87–97 (2002)
24. Victor, T.W., Harbluk, J.L., Engstrom, J.A.: Sensitivity of Eye-movement Measures to In-vehicle Task Difficulty. Transportation Research Part. F 8, 167–190 (2005)
25. Wann, J.P., Swapp, D.K.: Where Do We Look When We Steer and Does It Matter? Journal of Vision 1, 185 (2001)
26. Wickens, C.D.: Engineering Psychology and Human Performance, 2nd edn. HarperCollins, New York (1992)
27. Wickens, C.D., Lee, J.D., Liu, Y., Becker, S.E.: An Introduction to Human Factors Engineering, 2nd edn. Prentice Hall, New York (2004)
28. Williams, L.G.: Target Conspicuity and Visual Search. Human Factors 8, 80–92 (1966)

Searching for Possible Threat Items to Safe Air Travel: Human Error and Training

Xi Liu and Alastair Gale

Applied Vision Research Centre, Loughborough University, Loughborough, LE11 3UZ, UK
{x.liu3, a.g.gale}@lboro.ac.uk

Abstract. An eye-tracking experiment is reported which investigates the underlying factors that affect training in the visual search of air passenger luggage for possible threat items so as to reduce errors and improve safe air travel. In this study, naïve observers learned to search for terrorist threat items of guns, knives and improvised explosive devices (IEDs) in airport passengers' X-ray luggage images. During training, each participant viewed the same number of learning trials of guns, knives or IEDs. Transfer performance was measured in a same search task in which each participant was more familiar with the visual appearance of half of the test targets. Detection performance and eye movement data both showed improvement in the efficiency of search and recognition with practice, while the skills were stimulus-specific so that performance was degraded when novel targets were introduced. Perceptual learning and human errors of the implications for screener training are discussed.

Keywords: visual search, perceptual learning, airport X-ray luggage image.

1 Introduction

Airport security screeners regularly visually search X-ray images of passengers' luggage for potential threat items. A large number of funds have been invested for enhancing aviation security since the events of 9/11, including new screening equipment, technologies and security personnel training. To summarize the new security screening techniques are: dual-energy analysis for estimating the atomic numbers of objects in passenger bags; transmission scattering and computer tomography to separate objects in complex backgrounds; and scattered X-ray energy imaging techniques for detecting plastic explosives better [1]. Substantial enhancements are required to the existing techniques to meet reliable detection performance and acceptable speed. The report about aviation security (for the U.S. Government Accountability Office) shows that explosive detection systems had not been utilized for 100 percent of baggage because of the shortage of security screeners or some other reasons so that this task still relies heavily on the human interpretation of X-ray luggage images to detect terrorist threat items. However, this demanding task is accomplished in a few seconds and the images are difficult, having a low signal-noise ratio coupled with potential targets having unknown shapes in a cluttered

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 750–759, 2007.

background. Decisions of false-negative (FN) and false-positive (FP) are inevitable in examining X-ray luggage images for terrorist threat items due to the innately human limitations of visual cognition. Reasons for mistakes could be failures of expertise or a lack of expertise [2]. Efforts to update and maintain the expertise of security screening, overcoming the limitations of human perception and cognition, are therefore a priority for safer air travel.

An obvious difference between an expert and novice is that the expert has more experience of objects in their domain. In a related research domain to airport security screening, research has supported that the minimum number of cases for a radiologist to interpret in a fixed period is necessary [3]. It was indicated that one reason for the performance of residents to be worse than mammographers was the lack of visual recognition skills which could be obtained from perceptual learning with computer-assisted feedback. Massed practice with feedback is very helpful; not only for improving sensitivity but also for some non-specific aspects, such as conceptual knowledge or a search pattern which can transfer to other unfamiliar tasks [4]. For many tasks, human visual performance improves dramatically with practice while learning does not transfer when conditions changes. Karni and Sagi [5] have found that learning in texture discrimination tasks is local and transfer does not occur when background elements rotate 90°. In the task of discriminating the offset sign in a vernier discrimination task [6], learning does not transfer from horizontal orientation tasks to vertical orientation tasks or vice versa. For some situations, results have shown that a harder training phase induces learning which transfers to novel stimuli [7]; exposure of easy stimuli facilitates learning in difficult trials [8] and the ability of ignoring useless information from task-redundant information transfers to new stimuli [9]. In the task of searching for knives in passengers' X-ray luggage images, part transfer was observed from eye movement data as participants could fixate on targets more quickly after training [10].

The experiment reported here assessed the effect of practice with immediate object feedback and measured the transfer of perceptual learning in order to reveal how knowledge of security screening is organized. Participants practiced search and recognized threat items (guns, knives and improvised explosive devices (IEDs)) for three sessions in one day, which were displayed in a balanced order for each participant. Participants were tested on new stimuli with the same task following practice, and on the next day, so that transfer was measured by comparing performance on the trained and new stimuli. It was assumed that participants were more familiar with half of the test stimuli than the other half since half the test stimuli were scored with a higher similarity with practice stimuli than the other half in a pilot study, which gave a chance to evaluate the degree of transfer. Eye movements were recorded in this experiment for further analysing error reasons and revealing the mechanisms of performance improvement with practice.

2 Method

2.1 Participants

Twelve naïve people (6 female) took part in this experiment. Participants' visual acuity was normal or corrected-to-normal.

2.2 Stimuli

Experimental targets were eight guns from two viewpoints, eight knives and eight IEDs, which were all chosen from a large set of threat items. These items were scored for visual similarity (defined here as “similar objects share visual characteristics such as colour, texture, size, orientation and shape”) between objects of the same kind (e.g. a gun was only compared with a gun, not with a knife or an IED) in a pilot study. Four sets of two guns from two viewpoints, two knives and two IEDs each, termed Set 1, 2, 3 and 4, were derived so that the visual similarity within a set was scored higher than the visual similarity between sets. Moreover, there was high visual similarity between Set 1 and Set 3, and between Set 2 and Set 4; while low visual similarity existed between Set 1 and Set 2, and between Set 3 and Set 4.

Target-present images were generated by inserting target images into normal bag images randomly. There was only one threat item in each luggage image. The target objects of Set 1 and Set 2 were inserted into 60 normal bags for training purpose. Set 3 and Set 4 were inserted into 16 normal bags for test purposes. Target-absent images were 120 normal bag images without any threat items. Each session was composed of 16 target-present luggage images and 24 target-absent images.

2.3 Design and Procedure

The participants’ task was to search for threat items (gun, knife or IED) in passengers’ luggage images. Half of the participants were assigned to the group that learned targets from Set 1, and half were assigned to the group that learned targets from Set 2. On the first day, each participant completed three practice sessions (session 1, 2 and 3) and one test session (session 4). On the second day, the test session was done again (session 5). On each trail of training sessions, images were presented against a white background on a computer monitor for an unlimited time. Participants pressed a spacebar to indicate that they had finished searching and made their decisions using a five-point rating scale. If their decisions were higher than 2 (‘probably absent’), they also had to indicate the location of a potential threat item. Then an image in which the target was clearly displayed for target-present stimulus was followed for unlimited viewing. In the test sessions, other than no feedback provision the procedure of search and decision-making was the same as in the training sessions. There were three minutes break between training sessions and ten minutes break before testing on the first day.

In order to eliminate the effect of session order on measurement of detection performance, the order of practice sessions was counterbalanced so each session appeared equally at each stage and equally before and after every other session. Since half the test targets were from Set 3 and half were from Set 4, participants might be more familiar with half of the test targets than the other half but all test targets were novel to participants. The composition of an IED was explained to participants before the experimental sessions. Before the test session, participants were told the shapes of target objects in the test sessions were different from training targets and feedback images were not available.

X-ray luggage images were displayed on a 21-inch (53 cm) 1280 × 1024 monitor and viewed from a 70 cm distance. Eye position was calibrated before each session and eye movements were recorded using a Tobii X50 eye tracker.

3 Results

In this study, dependent variables were analysed by two-way mixed analyses of variance (ANOVA) with sessions as a within-subject factor and sets as a between-subject factor. All participants' data were pooled together as the differences among the sets were not investigated in this study. Location of a threat item was considered so that false location with a positive response was considered to be a false negative decision, only correct location and positive response was scored as a true positive decision.

3.1 Performance Analysis

Decision confidence data was analysed using the Receiver Operating Characteristic (ROC) method using the software ROCKIT [11]. The mean overall performance of each session was expressed as an A_z value, the areas under the ROC curve, which jointly considered hits and misses. Figure 1 is the graph with A_z of session 1 as the X-coordinate and A_z of session 3, 4 and 5 as the Y-coordinate. This shows the performance variation of participants in the last training session and transfer sessions. The line is a reference line to the first training session. Points in the upper area of the line showed better performance than session 1, while points in the lower area of the line showed worse performance than session 1. Figure 1 intuitively shows that the performance of session 3, the last training session, was much better than session 1 where all of the points are in the area of the upper line - performance increased with practice. About half of the points in transfer sessions were in the lower area of line which indicated transfer performance was even worse than session 1 when novel targets were introduced.

An analysis of variance (ANOVA) revealed that the improvement was significant with practice, $F(2, 20) = 23.829$, $p < .001$, reflecting an increase of the overall hit rate from .71 of the session 1 to .89 of the session 3 and a decrease of the false alarm rate from .33 of session 1 to .08 of session 3. Analysis showed that there was no difference of performance between session 1 and transfer sessions, while performance of session 3 was significantly better than transfer sessions, $F(2, 20) = 50.150$, $p < .001$ of session 4 and $F(2, 20) = 28.779$, $p < .001$ of session 5.

Figure 1 shows that the detection performance of some participants in the transfer sessions was worse than their performance in session 1. In order to investigate this interesting phenomenon, the hit rate in session 1 was divided into two parts since immediate feedback was provided and each threat item was displayed twice in different viewpoints and backgrounds in each practice session so that the detection performance would be enhanced due to the first presentation of targets. The hit rate of threat items in the first presentation (H_{first}) in session 1, similar targets ($H_{similar}$) and unfamiliar targets ($H_{unfamiliar}$) in test sessions were calculated separately (see Table 1). The $H_{similar}$ of session 4 and 5 were both better than H_{first} [not significant]. Hit rates decreased significantly as novel targets were introduced, $H_{unfamiliar}$ of session 4 and 5

were worse than H_{first}, F (1, 10) =68.820, p < .001, and F (1, 10) =30.179, p < .001 respectively. However, false alarm rates significantly decreased from .33 of session 1 to .13 of session 4, F (1, 10) =19.929, p = .001; and .14 of session 5, F (1, 10) =20.180, p = .001.

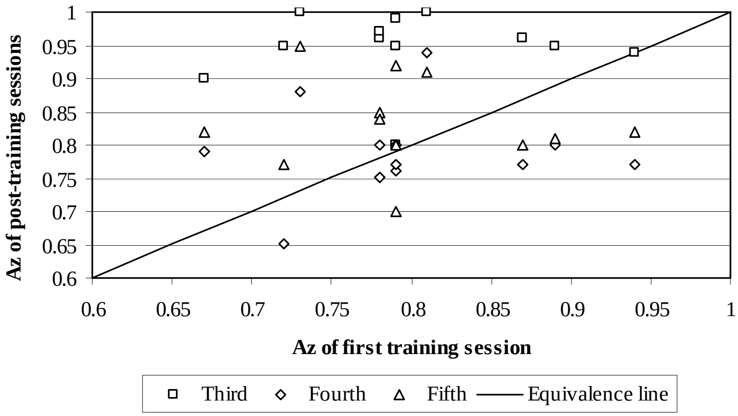


Fig. 1. Graph shows A_z values of participants in session 3, 4 and 5, as compared with session 1. The line is an equivalence line. Points in the upper area of the line represent performance better than session 1, while points in the lower area of the line represent performance worse than session 1.

3.2 Decision Time

Decision time decreased reliably with practice, F (2, 20) = 35.742, p < .001 of the target-absent images and F (2, 20) = 19.928, p < .001 of the target-present images. For transfer sessions, decision time of target-absent images were shorter than session 1 and longer than session 3, while decision time of target-present images was longer than all practice sessions. More details are presented elsewhere [12]. Decision time of false decisions was longer than that of true decisions for both target-present images and target-absent images in all training and transfer sessions (see Table 2), and differences in each session were significant, p < .05.

The decision time for the first presentation targets in session 1 was longer than that of similar targets in session 4 and session 5 (see Table 1), F (1, 10) =11.055, p = .001 and F (1, 10) =4.403, p < .05, respectively. There were no differences between decision times of first presentation targets in session 1 and decision times of unfamiliar targets in session 4 or session 5. Participants took more time to make decisions on unfamiliar targets than similar targets, F (1, 10) =18.107, p < .001 for session 4, and F (1, 10) =9.723, p = .002 for session 5.

3.3 Eye Movement Data Analysis

Eye movement data analysis further revealed mechanisms of performance change with practice and what happened in transfer. In consistence with the tendency of

decision time decreasing with practice, the fixation numbers on images decreased reliably from session 1 to session 3, $F(2, 20) = 14.782$, $p < .001$ of target-present images and $F(2, 20) = 8.169$, $p < .01$ of target-absent images. The fixation numbers on target-absent images were significantly more than that on target-present images, $p < .05$. Participants took more time and fixated on more places in target-absent images than target-present images.

Table 1. Mean performance and eye movement data for targets in session 1, 4 and 5

			AOI		
Target category	Hit rate	Decision time (ms)	Fixation number on AOI	Time to first enter AOI (ms)	Dwell time on AOI (ms)
Session 1					
First presentation targets	0.65	14410	10	2307	3343
Session 4					
Similar targets	0.68	10843	10	1183	3874
Unfamiliar targets	0.28	16326	13	1033	4806
Session 5					
Similar targets	0.70	11149	10	1127	3577
Unfamiliar targets	0.38	16494	12	943	4601

The area of interest (AOI) was defined to analyze how participants visually processed the target area [13]. With practice, participants were inclined to focus on a threat item area (AOI) quickly (see Table 2), $F(2, 20) = 4.432$, $p < .05$. In the mean time, participants took less time on the AOI, $F(2, 20) = 4.045$, $p < .05$. Also the fixation numbers in the AOIs decreased, $F(2, 20) = 3.298$, $p = .058$. Participants could fixate and recognize targets quickly after training. In all practice sessions, the fixation numbers on AOIs of FN responses were less than that of TP responses, $p < .05$; the dwell duration on AOIs of FN responses was shorter than that of TP responses, $p < .05$; the time to first enter an AOI of FN responses was longer than that of TP responses, but not significant (see Table 2). To summarise, participants fixated on potential target areas of FN decisions with less fixation points and shorter dwell duration than that of TP decisions in practice sessions. Although participants needed more time when they made a FN decision than a TP decision, they fixated on the AOI of TP decisions longer than on AOIs of FN decisions. Some features of targets of TP decisions attracted participants' attention.

In the transfer sessions, fixation numbers on target-present images were more than that on target-absent images, $F(1, 10) = 17.358$, $p = .002$ for session 4 and $F(1, 10) = 4.687$, $p < .05$ for session 5. The dwell time on AOIs of similar targets and unfamiliar targets in session 4 and 5 all were longer than that of the first presentation targets in session 1 (see Table 1), but not significantly so. The time to first enter an AOI in session 4 was shorter than that of the first presentation targets in session 1, $F(1, 10) =5.015$, $p < .05$ for similar targets and $F(1, 10) =5.677$, $p < .05$ for unfamiliar targets. Also the time to first enter an AOI in session 5 was shorter than that of the first presentation targets in

session 1, F (1, 10) =5.305, $p < .05$ for similar targets and F (1, 10) =5.584, $p < .05$ for unfamiliar targets. This indicated that the sensitivity to threat items was improved after training; no matter whether the decisions were correct or not.

Table 2. Decision time (millisecond), and eye movement data according to decision responses during training and transfer sessions

Decision time and eye movement data according to decision response	Training and transfer sessions				
	Session 1	Session 2	Session 3	Session 4	Session 5
Decision time					
False-negative (FN)	18914	8155	7847	14553	17352
True-positive (TP)	10945	6646	5323	12565	11427
False-positive (FP)	18459	16286	12495	16468	22436
True-negative (TN)	13491	8924	7356	9361	9436
Time to first enter AOI					
False-negative (FN)	2252	1322	1700	1112	1248
True-positive (TP)	1131	894	670	1011	854
Dwell time on AOI					
False-negative (FN)	2761	1848	1191	4657	3967
True-positive (TP)	3481	2779	2254	4894	4517
Fixation numbers on AOI					
False-negative (FN)	8	5	3	11	11
True-positive (TP)	10	8	7	13	12

3.4 Error Reasons and Skills Retention

The false-negative errors were classified into three categories: search error, recognition error and interpretation error [14]. If a target area was not fixated by any fixation points, then such miss responses were scored as a search error. If fixations or cumulative clusters hit on threat areas, and the gaze duration was less than 1000 ms, then these miss responses were termed a recognition error. If fixations or cumulative clusters hit on threat areas and the gaze duration was longer than 1000 ms, then miss responses were scored as an interpretation error. Other than session 3, the threat items were missed mainly due to interpretation errors (see Table 3).

Table 3. Missed errors in training and transfer sessions

Experiment sessions	Total number of missed error	Percentage of three types of missed errors (%)		
		Search error	Recognition error	Interpretation error
Session 1	54	15	11	74
Session 2	24	8	21	71
Session 3	22	9	50	41
Session 4	96	7	13	80
Session 5	88	8	11	81

The performance of session 5 was significantly better than that of session 4, $F(1, 10) = 9.181$, $p < .05$, reflecting an increase in overall hit rates from .48 to .54. There was no difference in false alarm rate and decision time between session 4 and session 5. Fatigue effects should be considered in session 4 which was completed after three practice sessions. However, these data still show that participants retained knowledge and skills very well; even better after one day.

4 Discussion

In this study, perceptual learning, transfer and error reasons were investigated in a simulated airport security screening task. Not surprisingly, hit rate, false alarm rate and reaction time improved with practice. Moreover, eye movement data revealed that participants could fixate on targets and process them more quickly after practice. Observers got some perceptual experience with the appearance of targets from immediate object feedback so that they could detect and recognize target objects more quickly and accurately in the following training sessions. However, the improvement was stimulus-specific which was not maintained when novel stimuli were introduced. Hit rate on similar targets declined to the level of session 1, while hit rate on unfamiliar targets was even worse than on session 1. The benefits of stimulus-specific learning were only evidenced as the sensitivity on targets in transfer sessions was higher than session 1: participants were faster to fixate on familiar and unfamiliar target areas.

Eye movement data analysis showed that most of the targets in transfer sessions were missed due to interpretation errors, which indicated that observers fixated on the correct locations for a long duration but they still did not recognize targets. The poor performance of naïve people was caused by their lack of expertise in the airport security screening domain. In the task of searching for terrorist threats in passengers' luggage images at airport, expertise of screeners includes knowledge of threat objects, generic knowledge about X-ray images, the ability to deal with time pressure, a high vigilance against any suspects, and so on. In the simulated task, only knowledge about X-ray images and threat items was required, which indicated that recognition ability was a result of the experienced accumulation of reading X-ray luggage images. This rule could be applied not only to novices but also to experienced screeners according to the possible error types [2]. The performance of experts will decline to novices' level when their repertoire of rules is exhausted for solving novel situations in a familiar task. The main difference between experts and novices lies in the skill-based level and the rule-based level such that experts arrange attention and apply skills more effectively than novices.

In a study of searching for knives in X-ray luggage images, McCarley and colleagues [10] thought that familiarity with stimuli and the task led observers to fixate target area sooner which was considered as a proof that practice might not improve search skills. In our study, search errors and interpretation errors decreased with practice showing that performance improvement was the result of gaining search skills and object knowledge. When new stimuli were introduced in the test sessions, the number of search errors in the test sessions was less than session 1 even if the number of misses in the test sessions was much more than that of session 1.

Moreover, the time to first enter the AOI of the test sessions was significantly shorter than that of session 1. Effectiveness of search was improved with practice and partly transferred to new stimuli.

In the training sessions participants consistently took a longer time and made more fixation points on target-absent images than on target-present images. A serial search model could be used to interpret this result which also appeared in the previous study [15]. When participants implemented the visual search task, they kept searching at a certain rate until the target was found so that search might be terminated in the middle of inspection. Otherwise, they would continue searching until every object in the image was scrutinized. The longer decision time of target-present images in the test sessions demonstrated that stimulus-specific with object feedback training helped observers develop perceptual sensitivity of threat objects but that the decision time was affected by unfamiliar targets. Moreover, general knowledge about the features of X-ray images obtained from practice was very helpful to reject distractors so that target-absent images were examined quickly in the test sessions.

In summary, frequent exposure of stimuli with immediate feedback in a real visual search task is an effective training method to integrate general knowledge of X-ray luggage images and recognition ability into perceptual experience. Learning in the visual search of threat items is stimuli specific, such that screener training should enlarge knowledge of terrorist threats. Practice decreased search error rates which improved the effectiveness of search. Nodine and Kundel [16] modelled visual scanning patterns in radiology which was composed of a rapid global scanning process and a systematic focal recognition process. This showed that there were differences in the visual scanning patterns of searching mammograms between experienced and inexperienced readers [17]. These results indicated that visual scanning patterns would be changed with training.

Further research work of ours will examine the effect of computer-assisted visual feedback training in the domain of X-ray luggage image inspection. We argue that human errors should be viewed not only in considering the individual screener's performance but also through taking the system approach [18]. An inspection error may happen even though high technology systems have many defensive layers.

Acknowledgements. This research is supported by EPSRC and carried out in collaboration with QinetiQ.

References

1. Muthukkumarasamy, V., Blumenstein, M., Jo, J., Green, S.: Intelligent Illicit Object Detection System for Enhanced Aviation Security. In: International Conference on Simulated Evolution and Learning (SEAL'04), Busan, Korea (2004)
2. Reason, J.T.: Human Error. Cambridge University Press, New York (1990)
3. Nodine, C.F., Kundel, H.L., Mello-Thoms, C., Weinstein, S.P., Orel, S.G., Sullivan, D.C., Conant, E.F.: How experience and training influence mammography expertise. Academic Radiologys 6, 575–585 (1999)
4. Sowden, P.T., Davies, I.R.L., Roling, P.: Perceptual learning of the detection of features in X-ray images: A functional role for improvements in adults' visual sensitivity? Journal of Experimental Psychology: Human Perception and Performance 26, 379–390 (2000)

5. Karni, A., Sagi, D.: Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. In: Proceedings of the National Academy of Sciences of the United States of America, vol. 88, pp. 4966–4970 (1991)
6. Poggio, T., Fahle, M., Edelman, S.: Fast Perceptual Learning in Visual Hyperacuity. Science 256, 1018–1021 (1992)
7. Doane, S.M., Alderton, D.L., Sohn, Y.W., Pellegrino, J.W.: Acquisition and transfer of skilled performance: Are visual discrimination skills stimulus specific? Journal of Experimental Psychology: Human Perception and Performance 22, 1218–1248 (1996)
8. Ahissar, M., Hochstein, S.: Task difficulty and the specificity of perceptual learning. Nature 387, 401–406 (1997)
9. Haider, H., Frensch, P.A.: The role of information reduction in skill acquisition. Cognitive Psychology 30, 304–337 (1996)
10. McCarley, J.S., Kramer, A.F., Wickens, C.D., Vidoni, E.D., Boot, W.R.: Visual skills in airport-security screening. Psychological Science 15, 302–306 (2004)
11. Metz, C.E., Herman, B.A., Shen, J-H.: Maximum-likelihood estimation of ROC curve from continuously-distributed data. Statistics in Medicine 17, 1033–1053 (1998)
12. Liu, X., Gale, A.G.: Search for threat items in X-ray luggage images: How visual skills and recognition ability develop with practice. In: Bust, P.D. (ed.) Contemporary Ergonomics, Taylor and Francis, London (In press)
13. Liu, X., Gale, A.G., Purdy, K., Song, T.: Is that a gun? The influence of features of bags and threat items on detection performance. In: Bust, P.D. (ed.) Contemporary Ergonomics, pp. 17–22. Taylor and Francis, London (2006)
14. Kundel, H.L., Nodine, C.F., Carmody, D.: Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. Investigative Radiology 13, 175–181 (1978)
15. Wolfe, J.M., Cave, K.R., Franzel, S.L.: Guided search: an alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human Perception and Performance 15, 419–433 (1989)
16. Nodine, C.F., Kundel, H.L.: The cognitive side of visual search. In: O'Regan, J.K., Levy-Schoen, A. (eds.) Eye movements: from physiology to cognition, pp. 573–582. Elsevier, Amsterdam (1987)
17. Krupinski, E.A.: Visual scanning patterns of radiologists searching mammograms. Academic Radiology 3, 137–144 (1996)
18. Reason, J.: Education and debate. Human error: models and management. British Medical Journal 320, 768–770 (2000)

Human Integration in the Lifecycle of Aviation Systems

Nick McDonald

Aerospace Psychology Research Group, Trinity College Dublin, Ireland
nick.mcdonald@tcd.ie

Abstract. While Human Factors is perhaps the most critical discipline to improving aviation safety, research and development is disproportionately small-scale, fragmented and unsustained. The key issue is the delivery of Human Factors knowledge throughout the system to improve design, operation or monitoring. A systems integration approach to technology development and innovation incorporates user requirements at all stages of the system life-cycle. The goal of the HILAS project is to develop and demonstrate such an integrated model of Human Factors research, practice and integrated application, linking design and operation – in a 'system life-cycle approach'. A central challenge is to demonstrate how to integrate models of the human operator, which demonstrate the influences on human performance, with wider system models that encompass the influences on system performance.

Keywords: aviation, Human Factors, safety, research capability, operational performance, system improvement, system life-cycle, innovation, system models.

1 Introduction

In a complex 'system of systems' like aviation, the human operator (pilot, cabin crew, ATC, maintenance technician) plays and will continue to play a critical role both within and between systems. The requirements of this role cannot be simply specified in a set of guidelines – as a recipe for 'human centred-design'. Human Factors has moved beyond analysing human fallibility and related performance deficits. It is increasingly addressing how people behave in normal operational contexts and how performance in such contexts can be better supported by design for use, by better planning and operational management and by quality and safety management systems. As new information technologies make possible the increasing integration of the 'systems of systems' of aviation, it becomes urgent to understand more comprehensively the human role in the system context. This inevitably extends the scope of what has traditionally been regarded as the domain of 'Human Factors'.

This requires an integrated approach, which systematically generates knowledge about the human aspects of the system at the operational end and transforms this 'knowledge about' into an active knowledge resource for more effective management and operational systems and better, more innovative, design. The challenge is to develop and demonstrate an integrated model of Human Factors research, practice

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 760–769, 2007.

and integrated application, linking design and operation – in a ‘system life-cycle approach’.

The overall goal of the HILAS project is to develop and evaluate such a model of Human Factors integration. The project contains four parallel strands of work that concern: integration and management of Human Factors knowledge; flight operations performance monitoring and process improvement; the Human Factors evaluation of new technologies applications on the flight deck, and the monitoring and assessment of maintenance operations.

This paper examines the background and rationale for the project.

2 Human Factors RTD Capacity and the Challenge of Aviation Safety

2.1 The Aviation Safety Challenge

In the European strategy for air transport set out in, “European Aeronautics: A Vision for 2020” a target of an 80 per cent reduction in aircraft accidents is proposed as necessary to support the expected growth in traffic with a reduction in the number of accidents. Following the publication of these targets the Advisory Council for Aeronautics Research in Europe (ACARE) produced a Strategic Research Agenda (SRA) which identified that ensuring effective and reliable human performance would be a key contribution to the required accident reduction.

Analysis of accident data has shown that for 70 per cent of aircraft accidents human error on the flight deck is cited as the primary cause. In a further 15 per cent of accidents human error on the flight deck is cited as a contributory cause. While it has been accepted for the last 30 years that Human Factors are perhaps the most critical discipline to improving aviation safety, the verifiable evidence that RTD in Human Factors has made a significant difference to aviation safety during this time is not strong. This is not to deny the considerable achievements of Human Factors in the introduction of crew resource management and in its increasing impact on the design of new technologies, for example. While these have had an impact, it is hard to quantify how much, and it undoubtedly the case that Human Factors remains the central area where verifiable progress has to be made if substantial gains in safety are to be achieved.

2.2 Weaknesses in RTD Capability

This lack of measurable impact can be attributed to several factors. Research and development in Human Factors is disproportionately small-scale, fragmented and unsustained in proportion to the scale of the problems that need to be addressed. The knowledge infrastructure is undeveloped for ensuring the availability of appropriate Human Factors knowledge precisely when and where it is needed to ensure its greatest impact. There has been a lack of learning by the industry from examples of good Human Factors implementation.

Thus, the following characteristics have generally been typical of Human Factors research in aviation:

- Fragmented in discrete, relatively isolated projects, rather than large scale integrated programmes.
- Sector-specific research and development predominates, rather than research which spans different aviation sectors (e.g. design, flight ops, ATM, maintenance) or which concerns the interfaces of these sectors.
- There is little integration of Human Factors research across the CADMID cycle (concept, assessment, development, manufacture, implementation, disposal).
- There is a serious lack of evaluation studies, particularly longitudinal studies of programme implementation.
- Competence in Human Factors is poorly developed and distributed around critical regions of the aviation community.
- There is little appreciation of what it takes to manage Human Factors knowledge and the activities it supports. Differences between Human Factors and technical knowledge are not generally well understood.
- There are few clear requirements standards. Most of these are at an operational level.
- Demands to manage and regulate complex issues concerning organisations and culture have not been matched by an appropriate research effort.

In summary, the implementation of existing Human Factors knowledge and methodologies is inconsistent, poorly monitored and evaluated and not transferred. Research has not been undertaken on a systemic basis across the sector and along the life cycle of systems. There is a poorly developed infrastructure to manage the effective generation, deployment and implementation of Human Factors knowledge across the aviation system.

2.3 Research and Development Needs

The JAA-FAST prioritisation of research needs have identified Human Factors and system change as strategic research priorities for aviation. If this priority is to be fulfilled with a commensurate impact on the aviation sector, then it will have to be based on an approach to Human Factors which is

- systemic, being fully integrated within each system component of the aviation sector (including technology, organisational systems and social processes),
- starts from design but extends throughout the operational lifecycle, including maintenance,
- addresses operational performance in a valid way,
- creates the basis for using Human Factors knowledge for improving technologies systems and processes.

All these requirements point to the need to address the way in which Human Factor knowledge is generated, distributed, implemented, and evaluated. The urgency of this task is emphasised by the projected growth of the aviation system over the next 20

years, with unprecedented demands for improving quality and safety while increasing capacity. Reconciling these goals will only be possible if the human and social contribution is addressed in a coherent and systemic manner.

3 Human Factors and System Change

Several developments of Human Factors research and practice point the way forward, but none on its own provides an adequate model.

3.1 Normal Operational Performance

LOSA (line operations safety audit), developed by the University of Texas in collaboration with NASA, represents a qualitative leap forward in providing a methodology to capture the normal processes of flight operational performance (2). This has led to the development of a ‘new view’ of human errors in which they are accepted as a normal and inevitable part of operational performance. Thus, what is critical to professional performance is both the capacity to recover from error and manage successfully the recurrent threats that are typical in normal operations. The ADAMS 2 project has demonstrated the feasibility of adapting this approach to aircraft maintenance operations.

While there is increasing evidence of LOSA as a performance monitoring methodology, there is little evidence about the impact of its use by airlines on the improvement of operational systems and procedures. Furthermore LOSA is administratively cumbersome, expensive and takes time to achieve feedback. This demonstrates the necessity to link performance assessment to a systematic feedback mechanism which can lead directly to system improvement. Such a mechanism does not exist at present, but achieving this is a specific goal of HILAS.

3.2 System Improvement

Process analysis and redesign methodologies are being adopted by airlines that are striving to reduce costs and improve operational performance. The basic principles of process analysis are well known (3) if not always well applied or properly validated. At the same time the introduction of information technologies in the form of ‘electronic flight bags’ or cockpit integration technology provides the opportunity to consolidate this approach into a ‘lean aircraft’ systems approach (4).

So far, in aviation, the development of process redesign methods and cockpit integration technologies has developed independently of performance management systems. Thus there are no models for a system of continuous improvement, which integrates these two approaches. In order to make this integration possible, it is necessary to develop analytic methods and identify requirements for the human characteristics of processes as well as performance, so that feed back from performance addresses the psychological realities of those processes which condition and influence performance. The ADAMS 2 project has been developing this approach in the context of aircraft maintenance processes (processes involved in maintenance checks, planning and supply chains, quality and safety management). The HILAS

project provides the opportunity to further develop this approach and extend it to flight operations.

3.3 The Lifecycle of Aviation Systems

Research on the integration of Human Factors across the lifecycle of systems has really only just begun. From an engineering design perspective, the demand for Human Factors commonly emerges as a request for a checklist or written guidelines which specify how to address the human requirement of new system design. From this perspective, Human Factors, no matter how early in the design cycle, can only play the role of an evaluation metric for an already existing idea. User-centred and participatory design processes have become more accepted and more formalised as processes. In these processes, representative groups of users contribute directly to the design process. When done well this can provide a much more in-depth appraisal of user needs.

However, this does not address the experience of actual users of the design, once it has been developed, manufactured and implemented. This kind of feedback, when systematically done, provides a much more robust basis for designing out the operational problems in the next generation of operational systems. The work on normal operational performance demonstrates the prevalence of informal, unofficial patterns of behaviour (often including routine ‘violations’ and non-compliance with procedures) which has been demonstrated in a variety of domains in aviation and elsewhere (5). Considering this, it becomes obvious that the kind of links with the customer that are being fostered by aircraft manufacturers (amongst others) are not a sufficient mechanism for capturing ecologically valid feedback from the reality of everyday operations. What is required is a much more structured system for routinely gathering such everyday information and making it available *in the appropriate form* to those responsible for current designs and future systems.

This is not a trivial task. The AMPOS project has demonstrated the difficulties of providing and using feedback to maintenance organisation and manufacturer for operational and procedural improvement. The ADAMS 2 project has developed a set of tools and methods for managing Human Factors at different stages of the aircraft maintenance lifecycle, from design to operational performance. However these remain to be integrated in a fully functional manner. HILAS will provide the opportunity to develop an empirically based model of the transformation processes necessary to make operationally derived knowledge usable at the design stage.

4 The Aviation System and the Management of Human Factors Knowledge

4.1 The ‘System of Systems’

Aviation is a complex ‘system of systems’ in which the human is the critical interface between the different sub-systems. Achieving the Strategic Research Agenda targets for European aviation will inevitably require redrawing the boundaries and roles between such subsystems – between flight deck and ground control of air traffic,

between maintenance and dispatch, for example. This requires an integrated 'system of systems' approach to research and development if dysfunctional interactions between poorly co-ordinated systems are to be avoided. Despite this, human factor research and development has been almost exclusively focused within each subsystem – flight operations, flight deck design, air traffic management, ground operations and maintenance. HILAS will provide the opportunity to develop and demonstrate a seamless approach to Human Factors integration across system boundaries. It will have a particular focus on the system boundaries between maintenance, dispatch and flight operations, and between the operational aspects of both maintenance and flight operations and technology and process design.

4.2 Knowledge Management

It has become obvious that the key to unlocking the potential of Human Factors to contribute to the robustness and error resistance of complex system does not reside solely in the domain of human-machine interface design (though this remains an important component). Rather, the issue is the delivery of the appropriate Human Factors knowledge (and the competence to use it) in the appropriate form, throughout the system to all those who are in a position to implement it to improve design, operation or monitoring. Currently there is an uneven distribution of Human Factors expertise between research institutes and universities on the one hand and industry (manufacturing and operation) on the other. The expertise deriving from research is imperfectly put in the service of design development and operation. Human Factors has been slow to develop a capability to thoroughly engage in the everyday realities of operational performance and organisational processes in an ecologically valid manner (the ADAMS 1&2 and AMPOS projects have pioneered this approach in aircraft maintenance). It therefore is necessary to learn from and apply some of the models of knowledge management, which have been developed in other domains.

It is important to recognise that Human Factors deals in tacit and implicit knowledge as well as explicit declarative knowledge, and that the pathways for managing Human Factors knowledge will often be different from those for 'technical' knowledge. The role of organisational memory and the capacity of organisations to learn from experience and put right the mistakes of the past are critical. Unfortunately the evidence suggests that organisations' capacity to learn in this way is very limited. It is clear that organisational culture plays an influential role in the development and institutionalisation of Human Factors in practice, yet very little research has seriously explored the parameters of this influence. The role of active social processes, like communities of practice, in fostering the development and interchange of knowledge are critical. Very little research has been done on the constraints and requirements for the sharing of knowledge between organisations despite the fact that many Human Factors initiatives have been founded on the premises that Human Factors is about safety and that the sharing of safety information is in everybody's interests (for example, GAIN, MEDA). Unfortunately these initiatives have not often delivered on their initial expectations, precisely because these constraints have not been understood. HILAS will explicitly address these issues and will develop a model of

good practice in the management of Human Factors knowledge in the complex system of aviation.

5 A New 'Business Model' for Human Factors

5.1 Human Factors as Technology

Human Factors is a unique science and technology, which systematically represents the user or operator of technical systems and processes, not just at the proverbial 'sharp end' but at all stages of the process. As such, more than any other discipline, it should be fully integrated into the system lifecycle. In aviation, Human Factors falls short of this ideal model and plays a very imperfect role in a systems integration model of innovation and development. The challenge for Human Factors is to develop a more comprehensive model, which can effectively integrate a variety of methodologies along the life cycle for example, from new technology/operational/product concepts through simulation and testing in their development to naturalistic research in operational environments and system modelling. This will require a greater interaction between research and development organisations (which can deliver applied research, technological research and basic research functions) within networked clusters of industrial organisations from design and development to operation and maintenance.

Through its workprogramme and through the user-support activities of the Knowledge Integration strand, HILAS will develop in practice the notion of an active innovation cluster of industry-research partnerships. This will work both on a European level and at a regional level, where synergistic partnerships can be formed between research and industry – for example in Ireland and the UK, in the Scandinavian region, in Italy and Spain.

5.2 Systems Integration Innovation

What is proposed is a systems integration approach to technology development and innovation, which would drive the incorporation of user requirements at all stages of the life-cycle of systems. What does this mean? The traditional business innovation model has a hierarchical linear sequence from basic research to developmental research to applied research to product development to marketing. Concurrent engineering and lean production have transformed this sequence by integrating design and manufacture. Design for manufacturability and continuous improvement driven by production requirements have driven down the time and cost of bringing new products to market. Lean production has led to agile production systems which flexibly meet customer needs. What is being proposed in the HILAS model is an analogous process of systematically building the user into not just the design process but also the whole life cycle of systems. In recent years there has been some growth in Human Factors capability in the aviation industry – amongst manufacturers, in ATM, in some airlines, maintenance organisations and national authorities. Thus, Human Factors do contribute to the development, operation, maintenance and regulation of aviation systems, but not in a coherent and integrated way across the life-cycle. Aircraft manufacturers are increasingly involving the user in the design process, but

this represents the first steps in what should be an integrated systemic process, addressing user needs across organisations. The great bulk of the research and development capacity in Human Factors is in the research institutes and universities. While the research institutes, in particular, have strong industry links, this is not well integrated along the system lifecycle.

Developing this ‘new business model’ will enable the better design and implementation of new technologies, which are developed with a sophisticated and valid appreciation of user requirements and characteristics. This will foster trust and acceptance in the introduction of new technologies. It will generate new and demonstrably valid methods and criteria for the certification and regulation of systems for human use. It will provide the tools and methods for the better management of quality and improvement of aviation operations and maintenance, which can lead to synergistic step improvements in both safety and efficiency. However, the fundamental argument is that this ‘new business model’ for integrating Human Factors in the lifecycle of European aviation represents a potential step change in competitiveness for the European aviation sector.

This analysis is based in part on Best (6), whose capabilities and innovation perspective (CIP) sees innovation and growth as being dependent on the production and organisational capabilities of clusters of organisations. For example, where the pull of the production system is towards innovation in new product development (NPD), “The NDP pull, interactive model seeks to permeate R&D throughout the organisation in a way that draws the customer / user into the definition of the problem and the solution”.

In the open systems model of innovation, technology integration teams “’dip-down’ into the scientific and technological bodies of knowledge that are available in the universities and ‘industrial districts’” in order to solve the challenges of rapid technical change. “Companies form long-term relationships with university research groups and other technology oriented firms to access [specialised knowledge and expertise]”. This contrasts with the hierarchical and linear model in which basic research in separate R&D labs drives technological development in a ‘trickle –down’ manner.

HILAS will adapt the open systems model of innovation to the large complex system of aviation where the user (the human operator), at all levels of the system, plays the key role in the different subsystems.

5.3 Modelling the Human in the System

If this vision of Human Factors as a driver of systems innovation is to be realised, then Human Factors has to develop the research capacity to play its role in understanding existing operational systems in a way which enables the human-centred design of future systems. Fundamental to this is the issue of modelling the role of humans in the system. If one wants to intervene in any way to change a system, one needs a model of that system which describes its underlying functionality and causal structure.

Models of ‘humans in the system’ can crudely be classified at different levels in terms of the extent to which they enable understanding and support intervention, as illustrated in Table 1, below. Many organisations manage Human Factors simply with

a set of checklists, and this is often what design engineers say they want from Human Factors. However the level of inference that such taxonomies support is very weak. Cognitive psychology has spawned many models of the human operator, either as an individual or in a small group, which can sometimes include tools as agents or actors. While such models can have great inferential power within their theoretical scope, they often do not address those factors which are critical to change if the operation is to be enabled to work better or designed to function more effectively in its environment. Therefore it is necessary to develop 'leverage' models which seek to address precisely these issues.

Table 1. Models of humans in the system

Level of model	**Characteristics modelled**	**Operational functions enabled**	**Design functions enabled**
Descriptive classification of Human Factors	Factors which potentially affect performance	Taxonomies for incident analysis, performance reports	Checklist for design support
Analytic model of human operator(s)	How Human Factors affect performance	Analyse / diagnose problems & events with respect to human operator	Evaluate HMI from user perspective
'Leverage' model of operational system	Functional relationships which support system outputs	Managing system & implementing change	Design and evaluate new system concepts

A model that provides leverage over the design and management of socio-technical systems has to be able to represent those factors that potentially causally influence the system's functioning. Most especially it should seek to model those factors that are amenable to modification, change or re-design in such a way as to transform the pattern of causal relationships that influences the required output of the system. This is perhaps the central theoretical challenge for the HILAS project – to demonstrate how to integrate models of the human operator, which are critical to understanding the influences on human performance, with wider system models which are critical to understanding the influences on system performance.

If this is possible, it will help to solve some fundamental problems of regulation. Not only will it provide the methodologies necessary to support certification of new technologies, from a human performance point of view, but it will also inform better models for the approval of airline and maintenance operators. If these two goals are separately possible, it makes realisable a third goal – to demonstrate a seamless human systems approach to both certification and operator approval. This integration is essential if we are to support the industrial goal of design for 'operability' – design to maximize the effective operational use of the technology to be manufactured. This,

of course, depends on being able to integrate cognitive models of the human operator with models of the human in the operational system. Thus system innovation to meet human needs requires radical transformation of the scope of Human Factors.

Acknowledgments. HILAS, ADAMS 1 & 2, and AMPOS are RTD projects funded by the European Commission.

References

1. Klinect, J., Murray, P., Merritt, A., Helmreich, R.: Line Operations Safety Audit (LOSA): Definition and operating characteristics. In: proceedings of the 12th International Symposium on Aviation Psychology, pp. 663–668. Dayton, OH: The Ohio State University (2003)
2. Rummler, G.A., Brache, A.P.: Improving Performance: how to manage the white space on the organisational chart. San Francisco: Jossey-Bass (1995)
3. Doganis, R.: The Airline Business in the Twenty-first Century. London: Routledge (2001)
4. McDonald, N., Corrigan, S., Daly, C., Cromie, S.: Safety management systems and safety culture in aircraft maintenance organisations. Safety Science 34, 151–176 (2000)
5. Best, M.: The New Competitive Advantage. Oxford University Press, Oxford (2001)

A Characteristic of a Navigator's Response to Artificial Ship's Movement by Picture and Motion Platform

Koji Murai[1], Tadatsugi Okazaki[2], and Yuji Hayashi[1]

[1] Kobe University, Faculty of Maritime Sciences,
Kobe, Japan
{murai, hayashi}@maritime.kobe-u.ac.jp
[2] National Maritime Research Institute, Navigation and System Engineering Department,
Mitaka, Tokyo, Japan
okaza@nmri.go.jp

Abstract. A navigator gets diverse navigational information from the ship and her environment for safe and efficient navigation, and visual information is the most important source for his judgments. We use ship's bridge simulator to do some education/training (training) of ship-handling and how to handle some navigational instruments. The navigational environment of the simulator is simulated with a picture. The visual image makes training effect improve in components of simulator's systems from the standpoint of the reality, and the realty has become more with computer high technology. Meanwhile, a real ship always moves with six kind degrees of freedom, and the ship's movement is simulated by the picture in merchant ship training, not the motion platform. A trainer doesn't evaluate the importance of the reality of the ship movement for simulator training yet. The purpose of this paper is to find characteristics of body response to artificial ship movement by the picture and the motion platform. We need to evaluate the simulation methods of the ship movement for good training.

Keywords: Ship bridge simulator, Training, Picture and Motion Platform, Body response.

1 Introduction

We use a ship bridge simulator (simulator) for a training and education (training) of a ship-handling of a navigator. An own ship and her navigational environment simulate with a picture. The reality of the image of her navigational environment is the most important to do some training of the ship-handling because the eighty percents navigational information comes from the visual of the navigator. However, the trainee in simulator training needs to clear the influence of the other information comes from four senses except the visual sensation, i.e. sound, gravity force. In this paper, we aim at the ship movement. The ship is a float and always moves on the sea. This factor is important for human for living on the ground like air-pilot training [1]. Moreover, a trainer doesn't evaluate the effects of the reality of the ship movement yet. The simulator training in Japan maritime university doesn't set up the motion platform, the

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 770–778, 2007.

trainee doesn't feel the gravity force of the ship movement in the training. The ship movement is simulated by the picture, not motion platform. We need to evaluate whether it is necessary to give gravity force from the ship movement to the navigator or not. The simulator for air-pilot gives the gravity force with a motion platform.

The purpose of this paper is to find characteristics of body and physiological response to the ship movement by visual and motion platform, to evaluate the simulation method of the ship movement. We select body gravity center, R-R interval and nasal temperature as indices of body and physiological response [2], [3], [4].

2 Simulator

We use the simulator, "Bridge Simulator for Navigational Risk Research", of National Maritime Research Institute (NMRI) is available for our aim which is to evaluate the mimic method of the ship movement- the comparison between the visual ship movement by the picture and the mechanical ship movement by the motion platform.

A visual system of the simulator produces a seascape of 240 degrees in horizontal view and 40 degrees in vertical view. It is enough specs to make a subject feel the ship shakes when visual ship's movement is produced. On the other hand, hydraulically-operated motion mechanism is installed under the floor of the bridge. This motion platform simulates roll (15 degrees) and pitch (15 degrees) motion of a ship. Moreover, the simulator has operator observation system and physiological data acquisition system. The operator observation system can record the subject's performance such as bridge monitoring system and eye track system which can measure the points what subject pays attention to during operation. The physiological data acquisition system can record operator's physiological data such as temperature of face and heart rate. These data allows for analysis of operator's state of metal-workload. We can easy to analyze the measured data with the simulator. Figure 1 shows the outline of the simulator consists of four main systems: Bridge system and Bridge motion platform system (Bridge simulator), Control system with Operator observation system, and Physiological data acquisition system. Bridge system sets up with real navigational instruments.

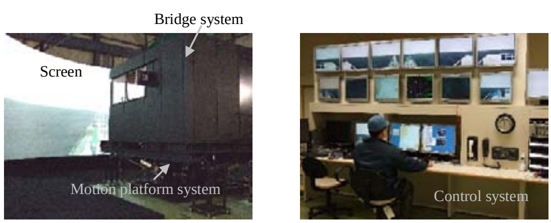


Fig. 1. The outline of the simulator consists of Bridge system, Motion platform system, Control system and Physiological data acquisition system

3 Experiment

The experiment has two bodies. The trainee is given the ship movement by the visual and the gravity force.

3.1 Outline

We carried out the experiment for eight kinds of rolling angles including zero degrees using the simulator of NMRI. The rolling and pitching are main factors of the ship movement [5]. We challenge to evaluate the effects on the training of the artificial rolling by picture (visual rolling) and motion platform (motion rolling). Table 1 shows the relationship among wave height, the rolling cycle and maximum rolling angle in order to simulate the eight conditions of the rolling by controlling picture and motion platform [6]. We find the characteristics of body and physiological response to eight kinds of the rolling while the navigator does lookout in open sea. The subjects, whom we select only male for differing the characteristics of the response between the male and the female [7], are six males in Table 2.

Table 1. Eight conditions of the rolling by controlling wave height and mechanical platform: wave height [meter], rolling cycle [second], and maximum rolling angle [degree]

Wave height [m]	Rolling cycle [sec]	Rolling angle (Max.) [deg]
0	0.0	0.0
1	10.0	0.1
2	10.0	0.4
3	10.0	0.8
4	11.0	1.5
5	11.0	3.0
6	12.0	4.0
7	14.0	6.0

Table 2. Six subjects' information; Age, Gender, Experienced year

Subject	Age	Gender	Experienced year
A	22	M	0.25
B	37	M	1
C	38	M	1
D	25	M	1
E	22	M	0.25
F	24	M	1

We measure three indices; 1) heart rate variability (R-R interval), 2) facial temperature, and 3) body gravity center from some indices. These indices show clear activity in navigation field from pre-experiment [3], [4]. Regarding three kinds of indices, we analyze 1) R-R interval measures with one milliseconds accuracy, 2) facial temperature including nasal and forehead part measures every 5.0 seconds with

the infrared camera, 3) body sway measures every 0.1 seconds with inspection unit of body gravity center.

3.2 Procedure

The experimental procedure is 1) Tester sets up the rolling angle by the picture or the motion platform. The rolling angle generates random maximum 0.1 to 6.0 degrees and tester inserts 0 degrees' rolling angle every time to check the influence of the rolling. Own ship's course is the south (180 degrees) and she receives a side wave from 090 degrees. The frequency of rolling is a constant for the all experiments (Table 1). The ship model is general cargo ship, 37,000 DWT, 15.2 knot (Navigation Full), 2) Subject stands on the inspection unit of balance function at the center position of the screen, and he looks the picture for one minutes. For the one minutes, he has the responsibility of safe navigation (Lookout). 3) This process is carried out for each rolling angle. Figure 2 shows the flow chart of the experiment.

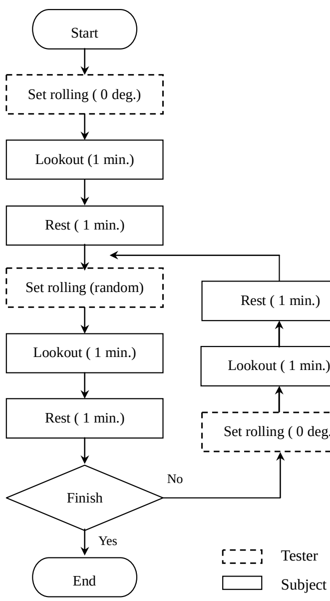


Fig. 2. Flow chart of the experiment

4 Evaluation

We evaluate the body and the physiological response to two types of mimic ship movement with three indices.

(1) Body gravity center: We calculate the width (min.-max.) of body sway back/forth (W_{FB}) and right/left (W_{RL}) of the body (Figure 3) and a directional coefficient (r) of the body sway (Equation 1).

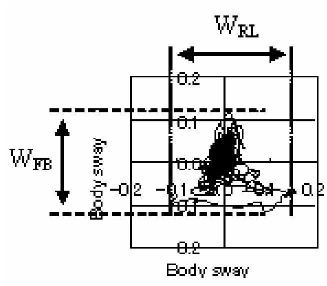
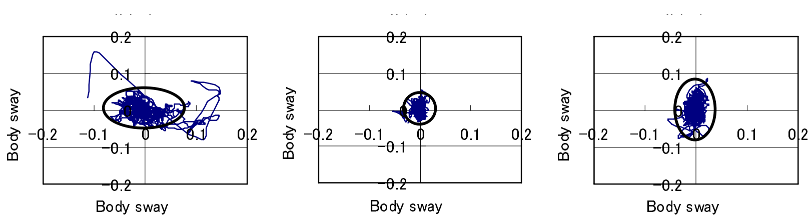


Fig. 3. Indices of the body sway, “width: W_{FB} and W_{RL}”

$$r = \tan (y_{std} / x_{std}) \quad (1)$$

here, x_{std} and y_{std} are standard deviation of the body sway right/left and in back/forth. The coefficient ‘r’ shows three cases of the body condition (Figure 4).

Case 1) R < 45 : oval with the major x-axis.
Case 2) R = 45 : circle.
Case 3) R > 45 : oval with the major y-axis.



(a) case 1 (b) case 2 (c) case 3

Fig. 4. Indices of the body sway, “directional coefficient: r”

(2) R-R interval: The R-R interval (Figure 5) is the heart rate for a moment. The R-R interval means the time interval from a peak point R wave to next peak point. The R wave is one of waves consists of P, QRS and T of an electrocardiogram (ECG). We calculate SNS value is frequency components of the R-R interval data [3]. SNS value, which is the index of the mental workload, is calculated by the Equation (2). Low Frequency (LF), 0.04 to 0.15 [Hz], is reflected by sympathetic nervous system and parasympathetic nervous system, and High Frequency (HF), 0.15 to 0.40 [Hz], is reflected by parasympathetic nervous system [8], [9]. This index is possible to evaluate sympathetic nervous system and parasympathetic nervous system at the same time. These frequency components calculate by Fast Fourier Transform (FFT) after interpolated with Cubic-Spline-Function every 1.0 seconds.

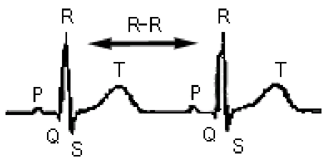


Fig. 5. R-R interval

$$SNS = LF / HF \tag{2}$$

(3) Nasal Temperature: We use thermal image camera to measure subject's nasal temperature. Figure 6 shows the example of thermal image. At the experiment, specifying subject's nose on thermal image, subject wears glasses.

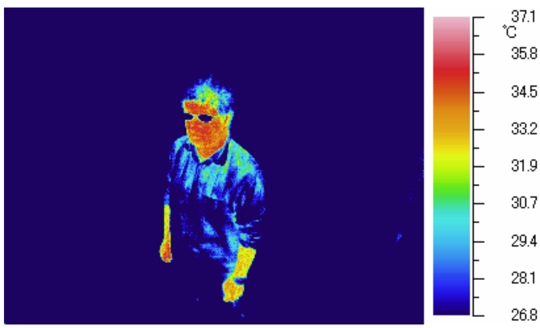


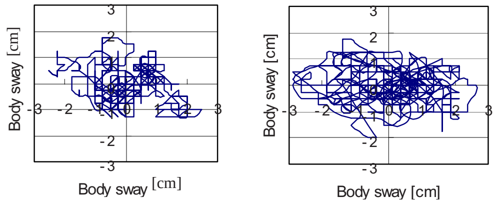
Fig. 6. Example of thermal image

5 Results

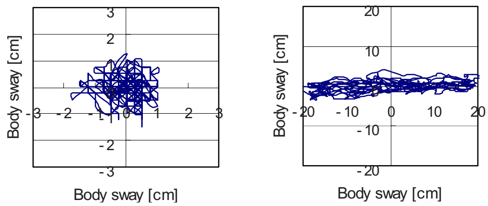
The results show that the characteristics of the body and physiological response to two kinds of the artificial rolling (ship movement). We evaluate the influence of the gravity motion and the visual effect of the ship movement for the simulator training.

Figure 7 and Table 3 show the typical results of the body sway of two artificial rolling, the maximum rolling angle is 0.0 and 6.0 degrees.

In Figure 7, the characteristic of subject's body sway differs in each experimental condition. His body sway is larger at more rolling angle. The values of coefficient 'r' decreases in comparison with 0.0 degrees for rolling angles, and W_{LR} and W_{FB} increase. We know two typical features between the visual rolling and the motion rolling. His body led toward an inclination of the picture for the visual rolling, but just kept the body balance to the inclination for the motion platform. The direction of the body sway differs. Then, the direction of the body sway for the motion rolling clear from Figure 7(b) and Table 3, r=6.4.



(a) Visual rolling (Left Fig.: max. 0.0 [deg], Right Fig.: max. 6.0 [deg])

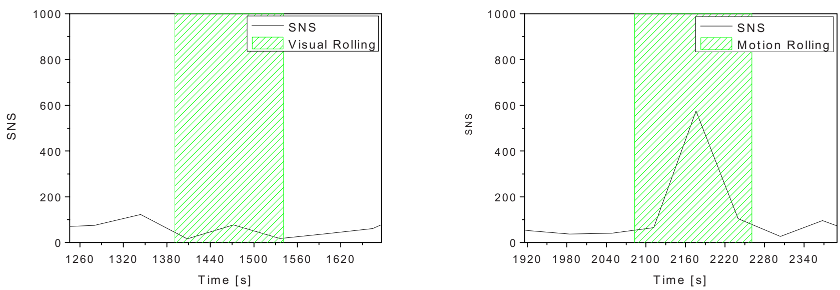


(b) Motion rolling (Left Fig.: max. 0.0 [deg], Right Fig.: max. 6.0 [deg])

Fig. 7. Body sway for visual rolling and motion rolling (subject B)

Table 3. The results of WFB, WLR, and coefficient 'r' in Fig.7

	Visual		Motion	
	Maximum rolling angle [deg]			
	0.0	6.0	0.0	6.0
W_{FB} [cm]	3.0	3.8	2.8	7.5
W_{LR} [cm]	5.0	5.5	2.8	40.3
r	36.9	30.0	41.6	6.4



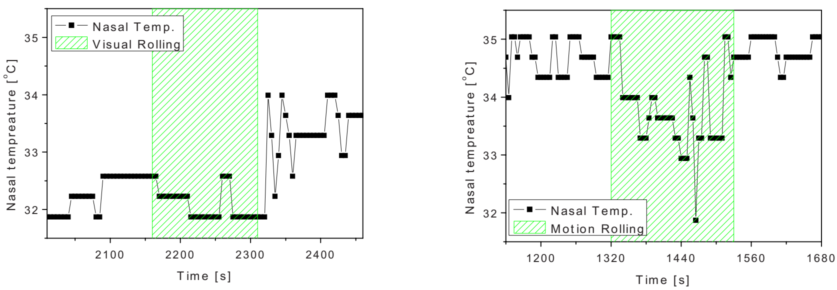
(a) Visual rolling (b) Motion rolling

Fig. 8. SNS values of maximum rolling angle 6.0 [deg] for visual rolling and motion rolling (subject D)

Figure 8 shows time series of SNS value which calculated from R-R interval data of subject D. The peak of SNS value shows that the subject felt strain.

In Figure 8(b), there is a peak of SNS value in cross-hatching part which is a period of motion rolling. On the other hand, there is no peak of SNS value at Figure 8(a). Therefore, it is clear that subject D felt strain at motion rolling.

Figure 9 shows time series of nasal temperature of subject A. The phenomenon of decreasing nasal temperature shows that the subject felt strain. In Figure 9(b), nasal temperature decreasing at the cross-hatching part which is a period of motion rolling. Therefore, it could be presumed that he felt strain during motion rolling. Meanwhile, in Figure 9(a), there is no remarkable decreasing of nasal temperature; however, after the visual rolling, subject's nasal temperature was increasing. It could be presumed that he relaxed after visual rolling.



(a) Visual rolling (b) Motion rolling

Fig. 9. Nasal temperature of maximum rolling angle 6.0 [deg] for visual rolling and motion rolling (subject A)

Our aim is which method is best for simulator training, or ship movement doesn't need for it. From the results, we need the ship movement for simulator training, because their body responds to the ship movement, and the body response differs clearly between two methods.

5 Conclusions

We carried out the experiment for finding the characteristic of body and physiological response to the ship movement 'rolling' with eight kinds of rolling angle including zero degrees. The response differs for the rolling methods "visual and motion", the both body sways are the same as it increase with more rolling angle and the motion rolling relates more than visual rolling.

In addition, a lot of phenomena of subject's strain were observed in rolling motion compared with visual motion.

Our future work is to clear the response while a navigator makes decision for ship-handling or fixed position etc. We thank the committee of *HCI 2007* and all anonymous referees.

References

1. Campbell, R.D., Bagshaw, M.: Human Performance and Limitations in Aviation, 3rd edn. Blackwell, Oxford (2002)
2. Murai, k., Hayashi, Y., Nagata, N., Inokuchi, S.: The Mental Workload of a Ship's Navigator using Heart Rate Variability. Int. J. of Interactive Technology and Smart Education 1(2), 127–133 (2004)
3. Kobayashi, H., Senda, S.: A Study on the Measurement of Human Mental Work-load in Ship Handling Using SNS Value. J. of Japan Institute of Navigation 98, 247–255 (1998)
4. Murai, K., Hayashi, Y., Inokuchi, S.: A Basic Study on the Body and Physiological Response to Visual Simulation of Ship Rolling. Trans. IEE of Japan 122-C (12), 2172–2179 (2002)
5. Kimura, N., Kawashima, R.: fundamental study for Human Dynamic Response to the Motion of Fishing Vessels in Ocean Wave -I. J. of Japan Institute of Navigation 69, 67–75 (1983)
6. Murai, K., Miyoshi, Y.: A Basic Study on the Relationship between Visual and Body Response by Ship Bridge Simulator. J. of Japan Institute of Navigation 103, 167–173 (2000)
7. Lentz, J.M., Collins, W.E.: Motion Sickness susceptibility and related behavioral Characteristics in men and women. Aviat. Space Environ. Med. 48(4), 316–322 (1977)
8. Perini, R., Veicsteinas, A.: Heart Rate Variability and Autonomic Activity at Rest and during Exercise in Various Physiological Conditions. European J. of Applied Physiology 90(3-4), 317–325 (2003)
9. Malik, M.: Heart Rate Variability. Circulation 93(5), 1043–1065 (1996)

Classification of Blink Waveforms Towards the Assessment of Driver's Arousal Level - An Approach for HMM Based Classification from Blinking Video Sequence

Yoshihiro Noguchi[1], Roongroj Nopsuwanchai[1], Mieko Ohsuga[2], and Yoshiyuki Kamakura[2]

[1] Information Technology Lab., AsahiKASEI Corp., AXT Main-tower 22F, 3050 Okaga, Atsugi-shi, Kanagawa 243-0021, Japan
[2] Biomedical Engneering, Osaka Institute of Technology, 5-16-1 Omiya, Asahi-ku, Osaka 535-8585, Japan
{nogu,roongroj}@ljk.ag.asahi-kasei.co.jp, {ohsuga@bme,kamakura@is}.oit.ac.jp

Abstract. With recent advances in image recognition, the assessment of driver's arousal level using blinking image sequences has been expected. In this paper, we demonstrated the possibility of assessing driver's arousal level by analyzing blinking image sequences. We focused on some typical blink waveform patterns occurred under drowsy condition. We used the results of EOG (Electro-occulogram) waveform clustering as the baseline for HMM (Hidden Markov Model) blinking labeling due to the difficulty of defining blinking labels from blinking image sequence. The blink pattern classes were classified by using the HMMs based on blinking image sequences. The driver's arousal level was finally estimated by histogram variation per minute of those typical blink classes.

Keywords: arousal level, blink, video sequence, EOG, HMM, driver.

1 Introduction

Recently, the investigation for an intelligent vehicle system has been developing based on a development of information technology. The information service and driving support depending on driver's state are expected. Among them, the assessment of driver's arousal level has been the challenge through the ages.

Driving with drowsiness is one of the main reasons causing traffic accidents. Recently, the assessment of driver's arousal level has been expected as an element of technology for establishing a safety transportation system. For one thing for the study of the assessment of driver's arousal level, there is sleepiness detection using blinks of driver. Heretofore, an eyelid closure time and a blinking period, which are relatively easy to be measured, have been focused attention on as the blinking characteristic for the assessment of driver's arousal level. However, the driver's arousal level while driving is not monotonous falling from high arousal level to low. The driver usually holds out against sleepiness while driving, so that we consider the same eyelid closure time and blinking period is occurred sometimes even though the driver's arousal level

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 779–786, 2007.

is different. Therefore it is difficult to assess the driver's arousal level only using these physiological characteristics, and it has not been in practical use yet.

The development of an in-vehicle camera system and the improvement of image processing technology make expectations of the assessment of driver's arousal level using blink image sequences higher. In this paper, we propose a new method for the assessment of driver's arousal level from blink image sequences using the histogram variation of the typical blink waveform class according to degrading of the arousal level. According to degrading of the arousal level, the histogram variation of the blink waveform class is changed. Therefore we consider a capability of the assessment of driver's arousal level using this histogram variation. We investigated the blink waveform type classification using HMM from blink image sequences as non-contact measurement.

2 Classification of Blink Waveforms

2.1 Blink Waveform Acquisition

Four subjects (3 males and 1 female) were participated in the present study. The subjects were able-bodied students of university, in the first half of the level at the age of 20. For the blink data acquisition not under the condition of monotonous falling from high arousal level to low but under the condition of against sleepiness while driving, the subject was seated on the driver seat of simple driving simulator system, as shown in Fig. 1, and drove for a maximum of 1 hour with a monotonous driving game.

The vertical EOG was derived using two electrodes, one placed above and the other below the right eye, a sampling rate of 250Hz, 5,000 times gain, with low-pass filter at the 35Hz. A video sequence with 640x480 pixels and 30fps of driver's eye blink was also shot simultaneously.

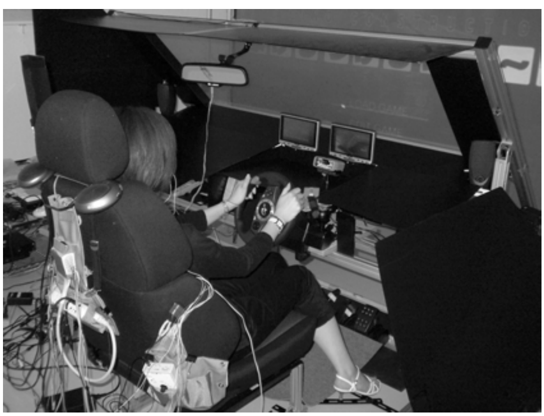


Fig. 1. Experimental setting

2.2 EOG Blink Waveform Classification

From many kinks of parameters of the EOG waveform features, we calculated correlation coefficients and checked the independence, then we chose three

parameters with a high degree of independence; peak height (blink amplitude), rising time (closing duration), and falling time (opening duration), as features of EOG waveform as shown in Fig. 2. We performed a cluster analysis of EOG waveform on each driving experiment of each subject.

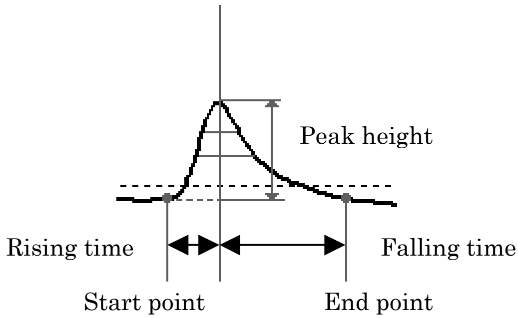


Fig. 2. Parameters of EOG waveform

In these three parameters, the peak height was used directly and the rising duration and falling duration were used after logarithmic transformation and Z scoring. We performed clustering using Ward's method with Euclidean distance. The number of cluster was twelve. We tried to eliminate a blink detection error as another cluster by clustering more than the commonly used number of blink cluster. As an example of EOG blink waveform classification, Fig. 3 is shown as an example of EOG blink waveform classification of one experimental set of one subject.

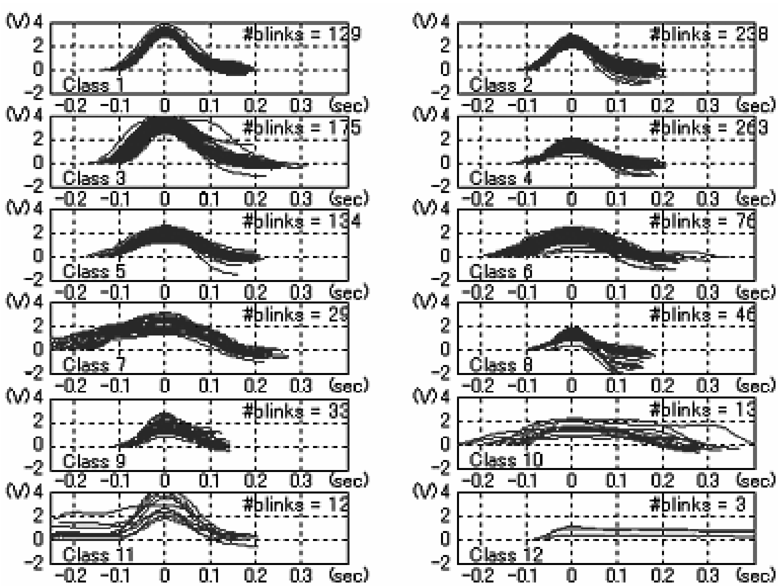


Fig. 3. Clustering example of EOG waveform

2.3 Assessment of Arousal Level form EOG Waveform Classification

Fig. 4 shows a sample of the 9-class classified blink waveforms and the histogram variation per minute of the blink waveform classes using EOG waveform while driving from high arousal level to low. We used only class 1 - 9 blink clusters, and did not use class 10 - 12 which were considered as blink detection error.

The subject felt light drowsiness from incipient period, and felt stronger drowsiness from last half. The subject then fell asleep immediately after end of experiment. It appeared that the blinks of class 2 on Fig. 3 were turned into small blink waveforms from standard waveforms of class 1 because of drowsiness. On the other hand, it appeared that the high peak height blink waveforms of class 3 were the deliberate blinks to hold out against sleepiness. The blinks of class 4 - class 7 were thought of as the blink waveforms under low arousal state, which occurred last half of the experiment. The histogram variation of the blink waveform classes according to degrading of the arousal level is perceived, therefore we consider a capability of the assessment of driver's arousal level using this histogram variation.

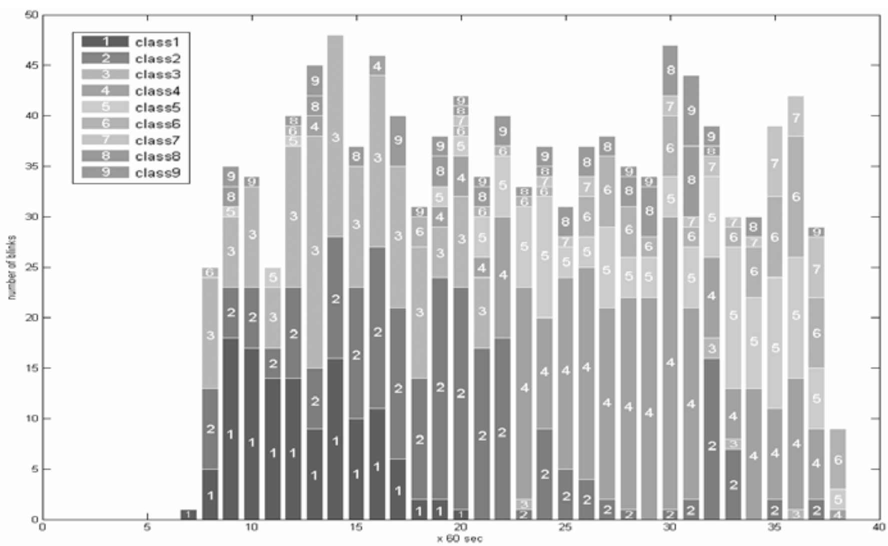


Fig. 4. Variation of histogram of EOG waveform

3 Classification of Blink Waveforms from Image Sequence

3.1 Blink Image Sequence

We investigated the blink waveform type classification from blink image sequences as non-contact measurement. Fig. 5 shows the example of blink image sequences of class 1 - class 4. Each blink image sequence was from start point to end point of blink, and the images framed by black line were at blink peak point. The challenge is the classification of blink waveform classes using HMM image recognition from blink

Fig. 5. Blink image sequences

image sequences whose frame rate is 30 fps, which is very slow compared to EOG sampling rate (250Hz).

3.2 HMM Blink Classification

We classified blink class from blink image sequence as shown in Fig. 5 using HMM image recognition. The blink class classification from blink image sequence mainly consisted of two processing parts. One was the feature extraction process from blink image sequence, and the other was image recognition process from extracted feature vectors.

3.2.1 Feature Extraction

As shown in Fig. 6, we used the vertical 30-dimension values whose each value was accumulated of 10 horizontal gray-scale pixel values in eye region rectangle (10x30 pixel) as image feature vectors. By using accumulation of pixel values along horizontal axis, we realized the robustness against the variation of horizontal face direction. We bundled these accumulated horizontal pixel value of image sequence from blink start frame to blink end frame, and used them as feature vectors.

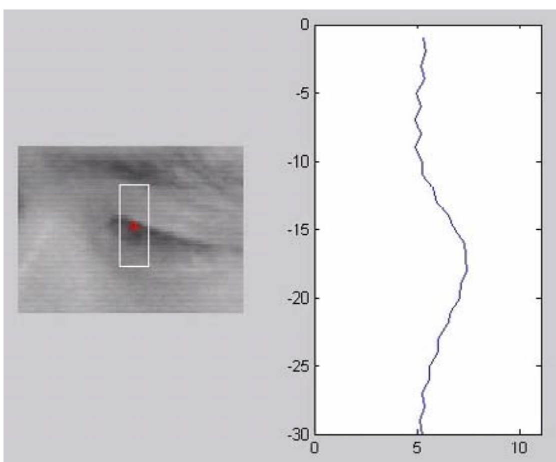


Fig. 6. Blink image feature

3.2.2 HMM Modeling

We chose a left-to-right HMM model which has 8 states for image recognition as shown in Fig. 7. S_i was a set of states, a_{ij} is a transition probability, and b_{ij} indicates an output probability of feature vectors through the corresponding state transition. This HMM consists of self-loops and sequential transitions between the current state and the next state. The model parameters are estimated to maximize the likelihood of the training data set. HMM model was prepared by training using the above-mentioned feature vectors of blink image sequences, which was defined label blink waveform class using EOG waveform.

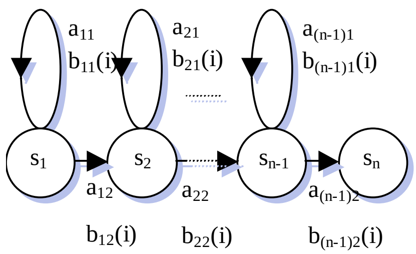


Fig. 7. HMM structure

3.2.3 HMM Blink Class Classification

We classified the blink waveform types using the prepared HMM model by inputting the feature vectors to HMM. The number of this subject's blink was 1,115 which covered class 1 – class 9. We used 2/3 blinks as training samples and remaining 1/3 blinks as testing samples on each blink class.

The classification accuracy of blink class was 95% within training samples and 46% within testing samples. Table 1 shows the detail of classification result on each blink class within testing samples. The classification error was not occurred for other classes equally, but was occurred just for similar EOG waveform classes as shown in Fig. 3.

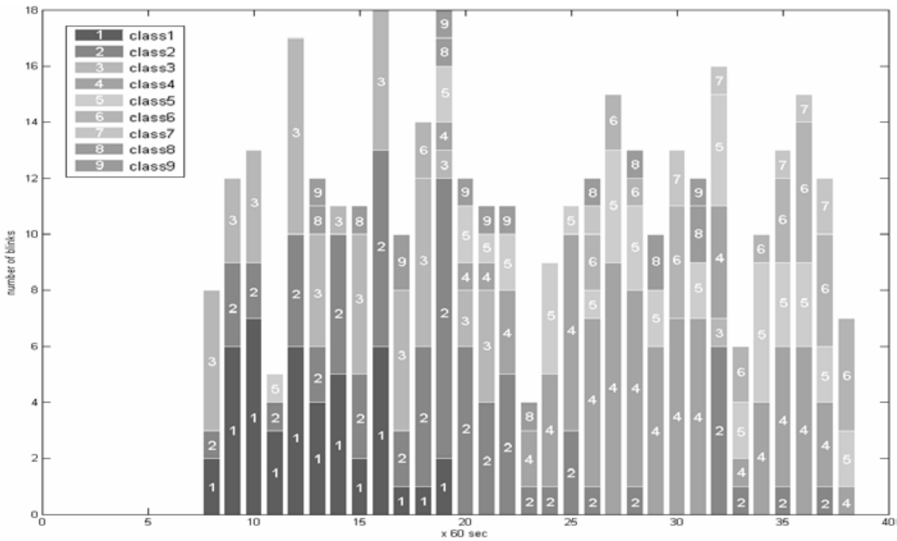
Table 1. Confusion matrix of blink classification result within testing samples

		Recognized class								
		C1	C2	C3	C4	C5	C6	C7	C8	C9
Original class	C1	16	7	21	0	0	0	0	1	0
	C2	7	39	17	8	4	0	0	0	1
	C3	4	17	30	1	0	0	0	0	1
	C4	0	8	0	58	11	4	0	2	2
	C5	0	6	1	22	14	2	0	0	1
	C6	0	1	0	16	7	2	3	1	0
	C7	0	1	0	5	0	1	0	0	1
	C8	0	2	1	5	1	0	0	1	0
	C9	0	0	0	1	0	0	0	0	7

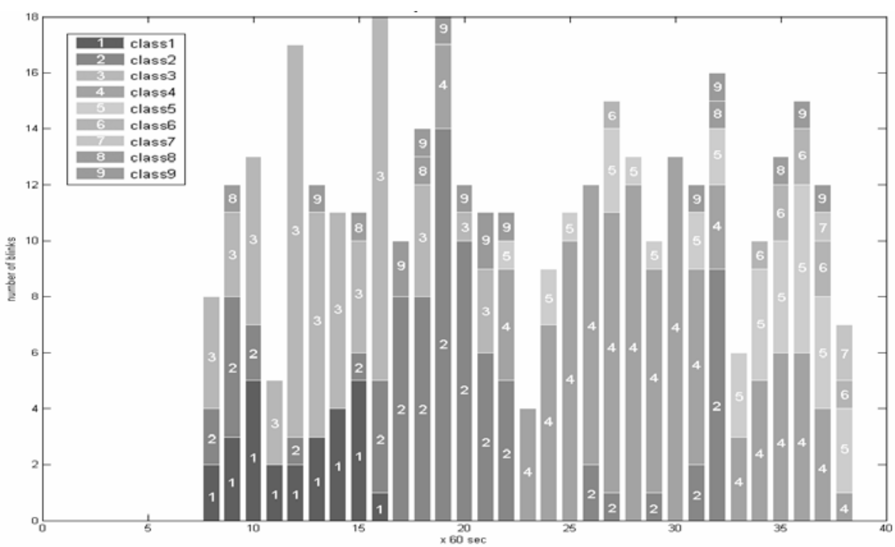
The histogram variation of the blink waveform classes from blink image sequence was shown in Fig. 8. We compared the variation of histogram of between EOG clustering and HMM image classification. Even though the classification accuracy

was only 46%, the histogram variation according to degrading of the arousal level can be perceived.

We assumed one of the main reasons of low blink classification accuracy within testing samples was the number of training sample was not enough. Furthermore, we were considering the integration of blink classes. The accuracy of blink classification could be increase in case the number of blink class was reduced.



EOG clustering



HMM classification

Fig. 8. Comparison between EOG clustering and HMM classification

4 Conclusions and Future Research

We investigated the blink waveform class classification using HMM from blink image sequences of driver as non-contact measurement. We consider a capability of the assessment of driver's arousal level from the variation of blink class histogram. In this experiment, the number of recognition samples was small. According to increase of samples, the consideration of the blink classification accuracy should be carried out. More study should be needed to establish the method to assess driver's arousal state using blink patterns.

References

1. Hamada, T., Ito, T., Adachi, K., Nakano, T., Yamamoto, S.: Detecting method for drivers' drowsiness applicable to individual features. Intelligent Transportation Systems 2, 1405–1410 (2003)
2. Miyakawa, T., Takano, H., Nakamura, K.: Development of non-contact real-time blink detection system for doze alarm. In: SICE 2004 Annual Conference, vol. 2, pp. 1626–1631 (2004)
3. Home, J.A., Reyner, L.A.: Driver Sleepiness. Sleep Monitoring, IEE Colloquium (1995)
4. Akerstedt, T.: Subjective and Objective Sleepiness in the Active Individual. Intern. J. Neuroscience 52, 29–37 (1990)
5. Ohsuga, M., Kamakura, Y., Inoue, Y., Noguchi, Y., Nopsuwanchai, R.: Classification of blink waveforms toward the assessment of driver's arousal levels - An EOG approach and the correlation with physiological measures. Intern. Conf. Human-Computer Interaction (2007)
6. Bekiaris, E.D., Nikolaou, S.I.: Towards the development of design guidelines handbook for driver hypovigilance detection and warning – the AWAKE approach. In: RTIC 2004, 12thIEE International Conference, pp.314–320 (2004)
7. http://www.awake-eu.org/
8. http://www.aide-eu.org/

Classification of Blink Waveforms Toward the Assessment of Driver's Arousal Levels - An EOG Approach and the Correlation with Physiological Measures

Mieko Ohsuga[1], Yoshiyuki Kamakura[1,2], Yumiko Inoue[2], Yoshihiro Noguchi[3], and Roongroj Nopsuwanchai[3]

[1] Faculty of Engineering, Osaka Institute of Technology. 5-16-1 Omiya, Asahi-ku, Osaka-city, Osaka 535-8585, Japan
[2] Faculty of Information Science and Technology, Osaka Institute of Technology. 1-79-1 Kitayama, Hirakata-city, Osaka 573-0196, Japan
[3] Information Technology Laboratory, Asahi Kasei Corporation. 3050 Okada, Atsugi-city, Kanagawa 243-0021, Japan
{ohsuga@bme,kamakura@is,yumiko@is}.oit.ac.jp,
{nogu,roongroj}@ljk.ag.asahi-kasei.co.jp

Abstract. The goal of this research is to develop a method to assess the arousal states using facial images of drivers. This paper refers the preparatory study on the classification of blink waveforms obtained from electo-oculogram. The transitions of the distribution of classified blinks during a simulated driving task were studied for around fifty volunteers of both genders and a wide range of generations. It was shown that the blink class ratio supposed to be under the influence of not only the subject's drowsiness levels but also by his/her behavior to battle with drowsiness. The correlation with multidimensional physiological indices was also discussed.

Keywords: arousal level, drowsiness, drivers, blinks, physiological indices.

1 Introduction

The assessment of driver's arousal levels is one of the most important issues for the development of adaptive driving support systems. Despite a long history of research in this field [1-7], the method to efficiently assess the arousal levels has not been realized. Particularly in the case of driving situation, the transition of the arousal levels does not simply change monotonically over time. The situation is rather more complex since the drivers tend to be compelled to overcome his/her drowsiness states. Physiological evidences of the drivers are consequently expected to show complex variations.

The development of the CCD camera embedded in the vehicle's inner mirror, together with a progress in image processing technologies, revived the expectation for a more powerful approach to assess drivers' arousal levels by observing eye-blink characteristics. In order to perform detail analysis of eye-blink characteristics at various

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 787–795, 2007.

arousal levels, this paper proposed a method to classify eye-blink waveforms obtained from the vertical EOG (electro-oculogram) and demonstrated the transition of blink patterns over the course of involuntary sleepiness. The correlation between the distribution of blink patterns and other physiological measures were also studied. By applying several electrodes over subjects' bodies, this method is considered as an intrusive method and is not suitable for a real driving situation. However, this research offers a novel insight into various relationships between eye-blink characteristics and drowsiness levels. It also serves as a significant contribution towards a non-intrusive drowsiness assessment system presented in [8], which uses computer vision techniques to determine drivers' eye-blink characteristics based on the video recorded from a CCD camera installed inside the vehicle.

2 Methods

2.1 Data Collection

2.1.1 Objectives

A simple and monotonous simulated driving task was set up. The objective of this experiment is to intentionally induce drowsiness of the drivers. Over the course of driving experiments, several measurements of the drivers were recorded for further analysis. These data include the physiological, subjective and behavioral data together with the video data of the driver's face.

2.1.2 Subjects

Fifty-nine paid volunteers of both genders aged between twenties and sixties participated in the experiment. Even though all subjects were licensed drivers, their driving experience varied significantly. The driving task and data measurement methods were explained, but the concrete aim of the experiment was not disclosed. After the briefing, each subject was asked to sign an informed consent form.

2.1.3 Experimental Tasks and Behavioral Measurements

Each subject was instructed to execute a simulated driving task for fifty minutes. A simulated driving environment was set up by using a commercial video game system projected on a 100-inch screen in front of the subject. A monotonous driving course, a circular circuit with little curve and no turn, was specifically created for this experiment. A comprehensive game controller with a steering wheel, an accelerator and a brake pedal was used for driving operation. Steering angles were measured by a potentiometer attached to the steering wheel. Two 7-inch LCD displays were installed at the dashboard, one for displaying a speedometer and the other for performing a simple visual reaction time task. The subjects were required to respond as quickly as possible, by pressing a small button on the steering wheel, when a circle displayed on the second LCD became larger, and the response times to the task were measured. Subjects' visual behaviors and facial expression were recorded by two CCD cameras. One of which was installed on a dashboard in front of the subject, and the other was attached at the position of the rear-view mirror. Figure 1 shows the experimental settings.

Fig. 1. Experimental settings. The upper left shows the driving image presented to the subject, the lower left is a rear view of the experimental system, and the two images in the right are subject's face images recorded by infrared CCD cameras.

2.1.4 Physiological Measurements

Multi-dimensional physiological recordings were carried out. These include the vertical and horizontal electro-oculogram (EOG), occipital midline electroencephalogram (EEG) recorded by using linked-ear references, skin conductance levels and responses (SCL, SCR), chest electrocardiogram (ECG) and respiration. The Ag/AgCl electrodes were used for EOG, EEG, SC, and disposable electrodes were used for ECG. Respiratory movement was measured using a carbon tube sensor attached around abdomen. These data were amplified and sent to a PC by a digital multi-channel amplifier for biological use (Polymate AP1132, TEAC). The time constants of the amplifier were set to 0.3 second for EEG, and 3.0 seconds for EOG and respiration, respectively. Skin conductance between two finger-tips was converted to voltage by an EDA unit (AP-U030) designed based on a circuit recommended by Society for Psychological Research. It was directed into the amplifier via a DC input port and sampled by 15 Hz high-cut filter. The previously mentioned behavioral data, steering angle, stimulus and response signals (for visual reaction time task), were also collected by the same device. The original sampling rate was 1 kHz, however physiological data was analyzed after re-sampled to 200 Hz.

2.1.5 Procedure

After attaching all the electrodes, obtaining the informed consent and a questionnaire on driving behaviors in daily life, the subject was brought into a sound-proof and dim room where the system for the experiments was set up. The experiment was divided into two parts. The first part was a 2-minute session where the subject was asked to be seated at the driving simulator and look at some fix points on the front screen with no restriction regarding their eye-blinks within this period. The data recorded in this part were used for calibration purpose. A short training session was allowed so that the subject became used to the driving operation. The second part of the experiment was the actual 50-minute driving task with simple reaction time task. The experiment can be aborted at anytime, especially when the subject complained about some uncomfortable symptoms such as motion sickness.

Subjective ratings on his/her fatigue, anxiety, discomfort and drowsiness were carried out soon before and after the second part of the experiment. After removing all

measurement apparatus, the subjects were answered orally to some questionnaire on the change in their subjective drowsiness and their efforts to stay awake during the driving task.

2.2 Data Analysis

2.2.1 Blink Parameters

We derived an algorithm for detecting eye-blink candidates from the vertical EOG data and their first and second differential waveforms. Two threshold values were pre-determined by observing the median amplitudes of the blinks recorded during the first part of the experiment of each subject. The higher threshold was used to detect eye-blink candidates and the lower one was used to decide the starting points and the end points of them occurred in the actual driving task. The candidates that had too long duration or too large displacement between the starting and the end points were rejected as non-blink waveforms.

Finally, the parameters that represent characteristics of each blink, i.e. the peak height (blink amplitude), the rising time (closing duration), the falling time (opening duration) and other five parameters, were extracted (Fig.2). The correlations among these parameters were studied, and only the first three parameters mentioned above were chosen to represent each blink due to their discriminative abilities. In this paper, blink waveforms in EOG associated with the downward eyelid movement are presented in positive change.

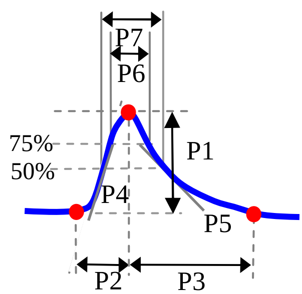


Fig. 2. Blink parameters. P1, P2, and P3 were used for classification.

2.2.2 Classification of Eye-Blinks

All blink parameters (P1, P2 and P3) were normalized by using a Z-scored normalization method. The classification of blink waveforms was then performed by K-means clustering. This method requires the number of blink classes (K) to be specified in advance. The classification was executed twice in order to reject false detections as many as possible. In the first classification stage, the number of blink classes was fixed to twelve. By visually inspecting the classification results, the classes consisting of non-blink artifacts were excluded from later analysis. In the second classification stage, the blink class number was reduced to eight, and the classification was carried out again.

2.2.3 Other Data

By applying the FFT spectral analysis, the average amplitudes of EEG components, i.e., theta (4-8 Hz), alpha (8-13 Hz) and beta (13-30 Hz) were obtained, The EEG component index was defined as the ratio of the amplitude of each component to the sum of those of three components. Instantaneous heart rate was calculated from R-R intervals, which were detected using R-wave enhancing filter. The discrete data was interpolated by a third spline method and re-sampled to 50 Hz. The trend in heart rate (HR) was extracted by low-pass filter with a cut-off frequency of 0.08 Hz. Mid-frequency component of heart rate variability (HRV-MF) and high-frequency component (HRV-HF) were obtained using band-pass filters with 0.08-0.12 Hz and 0.12-0.5 Hz, respectively. For the present, SC (SCL and SCR), respiration and the steering angles were observed in raw wave forms. Reaction time (RT) was measured between the upward strokes of task and response signals. If no response signal was found during 2 seconds from the task signal, a mark designated the omission of the response was inserted at the task timing.

3 Results

3.1 General Results

Seven out of fifty-nine subjects aborted the experiments due to their uncomfortable symptoms resembled motion sickness. Thirteen subjects did not experience drowsiness over the course of fifty-minute driving, while the remaining thirty-nine subjects reported some changes in their arousal levels. The subjects who did not experience drowsiness were commonly found in the elderly aged between fifties and sixties. Based on the subjective report of all subjects, various types of the temporal changes in subjective drowsiness were reported. These include a gradual increase in drowsiness levels, a peak of drowsiness levels in the middle, or a repeated fluctuation of drowsiness levels. However, the drowsiness levels derived from our performance measurement were not always consistent with the drowsiness levels given in the subjective reports.

Figures 3 and 4 show the analysis of all physiological data of two subjects (Subjects A and B). Following is the explanation of these results.

3.2 Blink Classification

Examples of the blink classification results of Subjects A and B are shown in the left columns of Figures 3 and 4, respectively. The EOG waveforms of all blinks in each class overlaid on top of each other, centered around their peaks, are shown in Figs. 3(a) and 4(a). The median values of the start point, the peak height and the end point are also presented in the plots as the red marks. Figs 3(b) and 4(b) depict the number of blinks in each blink class occurred at each time instance over a fifty-minute driving task. The temporal changes in the frequency of the blinks belonging to a certain class can be easily observed from these plots. Note that the vertical axes of these plots are shown at different scales. In Figs 3(c) and 4(c), the blink histograms of each blink class are stacked at each minute, so that both the alteration of dominant classes and the change in the total number of blinks can be easily observed.

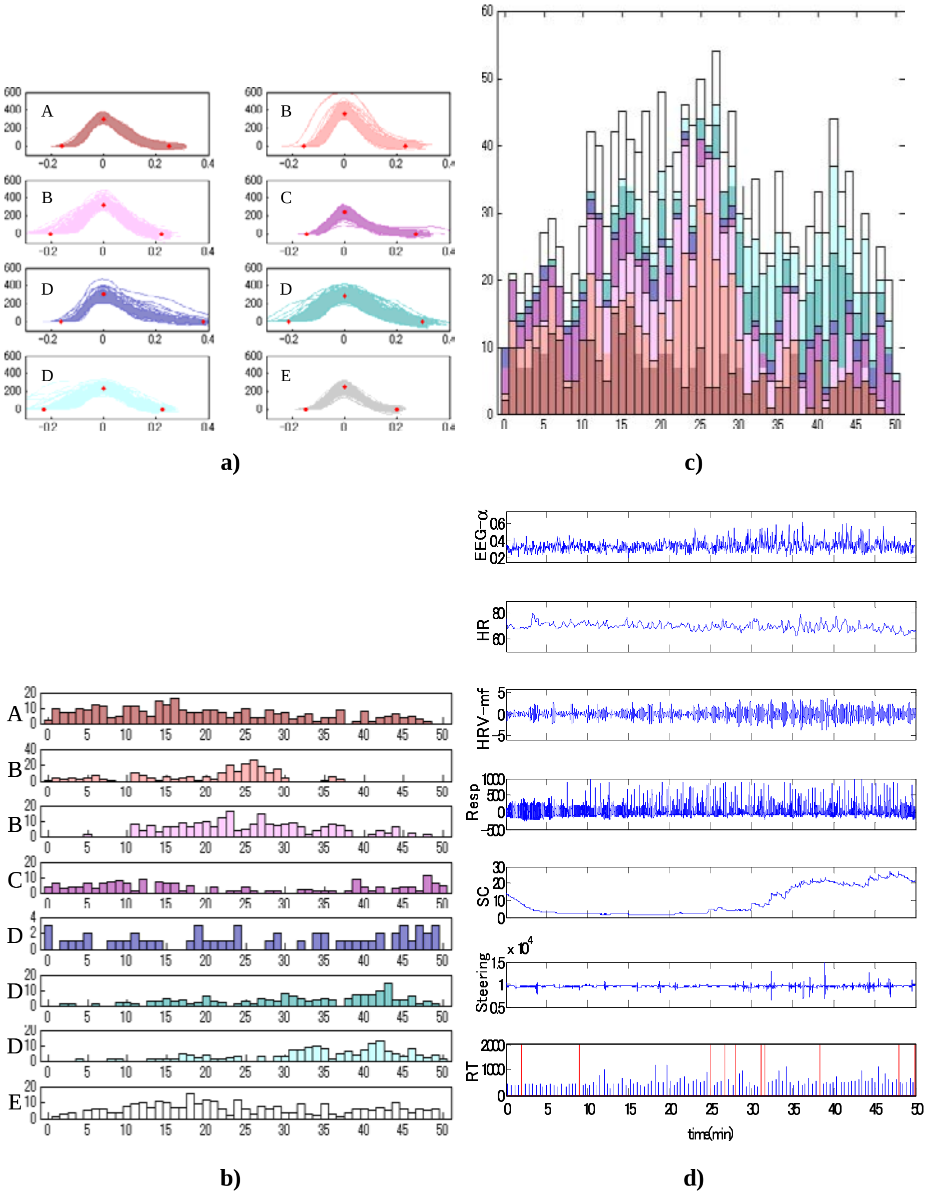


Fig. 3. An example of results (Subject A) a) overlaid EOG-waveforms (horizontal axis is time in sec), b) temporal distribution of blinks belonging to each class, c) number of blinks for all classes stacked at each minute, d) changes in physiological measures (EEG-alpha, HR, HRV-MF, SC) and those in behavioral measures (steering and reaction time). Omissions in response are shown by red lines. The horizontal axes in b)-d) are time in minute. See text for details.

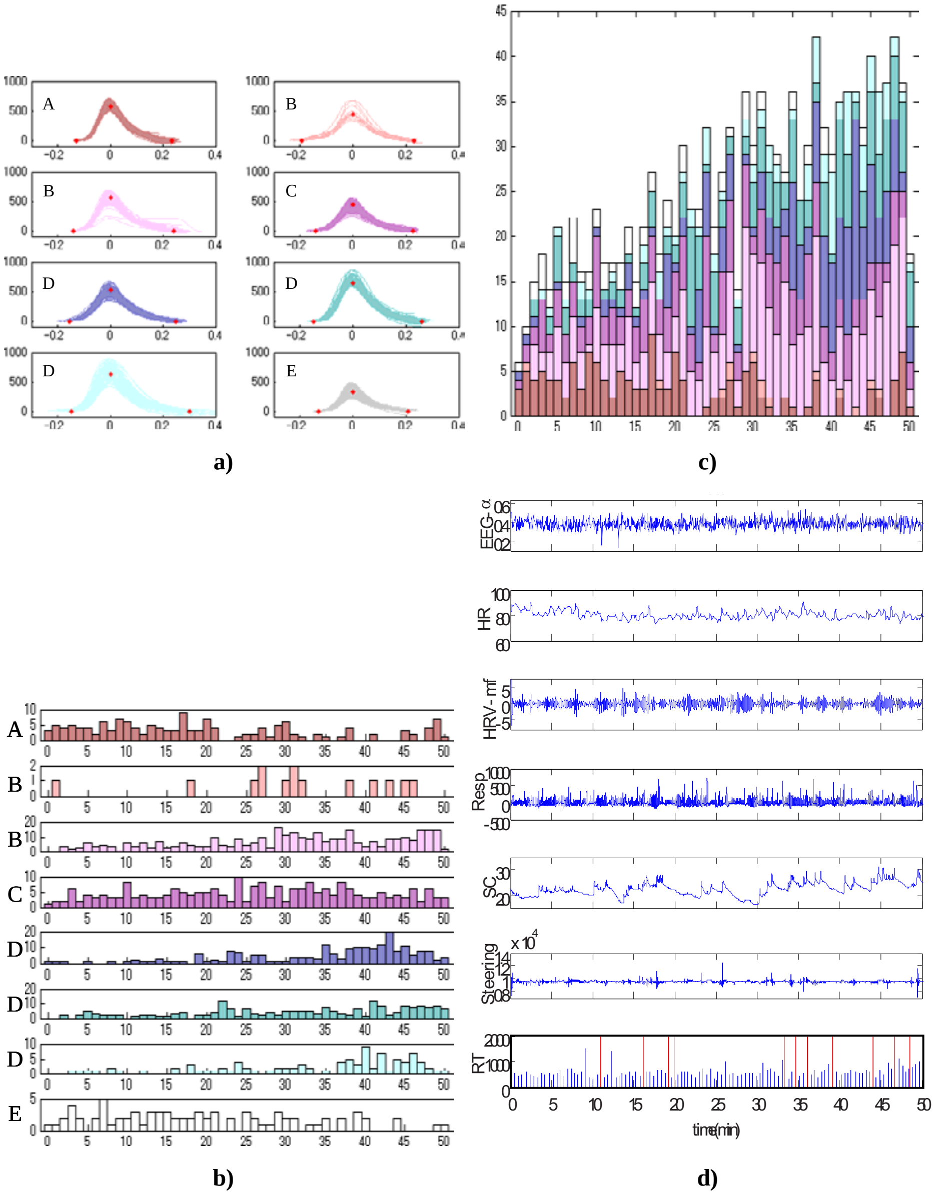


Fig. 4. An example of results (Subject B) a) overlaid EOG-waveforms (horizontal axis is time in sec), b) temporal distribution of blinks belonging to each class, c) number of blinks for all classes stacked at each minute, d) changes in physiological measures (EEG-alpha, HR, HRV-MF, SC) and those in behavioral measures (steering and reaction time). Omissions in response are shown by red lines. The horizontal axes in b)-d) are time in minute. See text for details.

The categories of blink classes were defined and were represented by capital letters A to E. These categories were determined as follows. By observing the EOG waveforms of the blinks in each class and their temporal distribution, a category which represented *normal* (awake) blinks (Category A) was defined and was located at the top of the figure. The waveforms of the *normal* blinks tend to be large and sharp, i.e. the peak height parameter is large and the other two parameters in the time domain are small.

The constitution of the classes was varied subject by subject, however, similar characteristics were found in common. Categories B to E were defined as follows. The classes with high peak amplitudes and long durations of eye-closing were commonly shown and named by Category B, whereas the classes with low peak amplitudes were grouped into Category C. Category D represented relatively long durations for eye-opening and/or eye-closing. In addition, the classes with very low peak amplitudes and small durations (Category E) were also observed. Temporal distribution of the occurrence of each blink category was considered to be dependent on the change in the arousal states of each subject. For an example in Fig.3, the blinks of Category C appeared soon after the beginning of the driving task while those of Category B increased in the middle, and those of Category D were observed over the late phase of the driving task. Category E was widely distributed.

By considering the drowsiness levels given in the subjective reports and the drowsiness levels derived from our performance measurement, some small discrepancies were commonly observed. However, by comparing the results of the blink classification with the recorded video images, the preliminary hypothesis can be obtained as follows. The blink classes in Categories C and D are expressions of low arousal levels, while Category B is a sign of subject's effort to maintain or raise the arousal levels. Most of the blinks in Category E are composed of a series of blinks with short intervals, which may occur either by drowsiness or by eye fatigue.

3.3 Correlation with Other Physiological Measures

In Fig.3(d) and 4(d), some selected results of other physiological measurements are presented. As shown in these figures, an increase in EEG alpha accompanied by a decrease in EEG beta and SC was often found at the beginning of the experiment. In a few subjects, EEG alpha increased in the middle or late phase of the experiment when the arousal levels were deemed to decline as estimated by blink patterns, performance measures, and/or subjective reports. No remarkable change was found in an EEG theta component. In most subjects, the mean HR decreased gradually and monotonously, hence not reflecting the detailed changes in the arousal states. SCL did not show a simple correlation with the arousal level. In some subjects, SCL declined smoothly mostly in the early phase of the experiment, however, the elevation with the repeated sharp rises and gradual declines were often observed over the experiment, especially in the middle and the late phase. A smooth decline in SCL often corresponded to the blinks in Category C, and the SCL elevation accompanied by fluent SCRs tended to correspond to a decrease in regularity of respiration and an

increase in HRV-MF. During this state, the blinks in Category B had a tendency to increase.

4 Conclusion

Over the course of fifty-minute driving of fifty-two subjects, the temporal changes in the number of each blink class were studied, and it was found that dominant blink classes changed as time passed. These changes in the distribution of blink class ratios were supposedly under the influence of not only the subject's drowsiness levels but also by his/her behavior to battle with drowsiness.

More study should be carried out to establish the method to assess driver's arousal state using blink patterns, verifying the relationship between the distributions of the blink class ratio and the drowsiness states based on the multi-dimensional physiological and behavioral measurements. In addition, the results in EOG approach would be applied to improve a non-intrusive method to detect eye-blinks and to classify blinks based on the facial video images of the drivers recorded from CCD cameras which has also been studied [8].

Acknowledgements. We thank all participants for their cooperation. We also appreciate the helps of Shota Hashimoto, Akari Kakimi and Koji Sakai for their contributions in executing experiments and data analysis.

References

1. Boadle, J.: Vigilance and simulated night droving. Ergonomics 19, 217–225 (1976)
2. O'Hanlon, J.F., et al.: Comparison of performance and physiological changes between drivers who perform well and poorly during prolonged vehicular operation. NATO Conference Series III 3, 87–109 (1977)
3. Keckluind, G., et al.: Sleepiness in long distance truck driving: An ambulatory EEG study of night driving. Ergonomics 36, 1007–1017 (1993)
4. Wierwille, W.W., et al.: Research on vehicle-based driver status/performance monitoring; development, validation, and refinement of algorisms for detection of driver drowsiness. U.S. Department of Transp, National Highway Traffic Safety Administration. DOT HS 808 247, Final Report (1994)
5. Bittner, R., et al.: Fatigue indicators of drowsy drivers based on analysis of physiological signals. Ergonomics. In: Crespo, V., et al. (ed.) ISMDA 2001. LNCS, vol. 2199, pp. 62–68. Springer, Heidelberg (2001)
6. http://www.awake-eu.org/
7. http://www.aide-eu.org/
8. Niguchi, Y., et al.: Classification of blink waveforms towards the assessment of driver's arousal level - An approach for HMM based classification from blinking video sequence -. HCII2007 (2007)

Common Work Space or How to Support Cooperative Activities Between Human Operators: Application to Fighter Aircraft

Marie-Pierre Pacaux-Lemoine and Serge Debernard

Laboratoire d'Automatique et de Mécanique Industrielles et Humaines - UMR CNRS 8530
Université de Valenciennes et du Hainaut Cambrésis
Le Mont Houy,
F-59313 Valenciennes Cedex 9, France
{marie-pierre.lemoine,serge.debernard}@univ-valenciennes.fr

Abstract. In order to improve capacities and capabilities of human-machine system, interactions between agents are inclined to be more frequent and more complex. Nevertheless, these interactions can not be always developed by verbal activities and must be supported by a media. The contribution of this paper is to propose the concept of common work space which allows agents, human or artificial, to develop cooperative activities. This concept has been implemented and evaluated in the fields of the fighter aircraft.

Keywords: Human-Machine Cooperation, Common Work Space, Simulation.

1 Introduction

Human-Machine Cooperation was one of the main studies of the team Human-Machine System of our laboratory since long time. Firstly applied in manufactory process, the next privileged domain is the Air Traffic Control. A pluridisciplinary team which gathers engineers from automation and psychology aims at examining the definition and the place of an automatic system able to detect and resolve air traffic conflicts [1]. Three main points can be extracted from hypothesis tested during several experiments conducted with experienced air traffic controllers: team activity can be expressed with a multi level organization, tactical and strategic levels were the two levels studied during experiments. Assistance tools must be defined to help human operator of the same level of activity. The assistance tool of the upper level (more strategic) must have a good model of what agents of lower level (more tactical) are able to perceive, to decide, to do, in order to control the process or to communicate with partners. This result leads to the third point which is the necessity to build a work space, common to agents implied in the control of the process.

When the air traffic control studies focus on the development of assistance tool on the tactical level [1], the present paper aims at providing assistance tool on the strategic level of a two-seater fighter aircraft organization.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 796–805, 2007.

So, the first part of the paper presents several definitions which appear to be important in the understanding of the role of each operator and assistance tool. It is mainly supported by their/its *know-how*, the manner they can communicate, and about what, thanks to their/its *know-how-to-cooperate* and through space dedicated to this cooperative activity, the *common work space*. The second part gives an application of these concepts in the domain of fighter aircraft.

2 Common Work Space

Human operator activity was usually the start point of cooperative activity study. The activity of each operator must be identified and analyzed. So the *know-how* was defined. It is the cleverness to solve problem. An agent has the abilities to build know-how thanks to knowledge and experiences. According to Rasmussen’s model, it is composed of four main activities [2], information elaboration, diagnosis, decision making, and action. These classes of activity can be decomposed into sub-classes describing the evolution of the state of information. *Information Elaboration:* It is the perception of information, provided directly or by captors on the process. *Diagnosis:* It is identification or inference or testing. The identification is the interpretation of information into a categorised one. Inference is also an interpretation, but it is uncertain. Testing is the test of an inference with new information. *Decision Making*: A decision is schematic or precise and can be evaluated. A Schematic Decision specifies a goal to be reached. A Precise Decision is used when the decision is fully specified. A Decision Evaluation corresponds to the control or evaluation of one's own action or another agent's action. *Action* aims at controlling the process using commands.

After the know-how which allows individual agent activity analysis, we need a method for identifying and analyzing the co-operative activity, so the *know-how-to-cooperate* was defined [3]. From the psychology point of view, Hoc [2] gives a definition of cooperation according to two concepts:

- Interference management*: Each one strives towards goals and can interfere with the other on goals, resources, procedure, etc.*

It would appear when two agents have to discuss a difference in their understanding of process evolution. For Castelfranchi, a “Common World” implies that there is interference among actions and goals of agents: the effects of the action of one agent are relevant for the goals of another, i.e. they either favour the achievement or maintenance of some goals of others (positive interference), or threat some of them (negative interference) [4].

- Facilitation*: Each one tries to manage the interference to facilitate the individual activities and/or the common task when it exists.*

It is the action to make easier the activity of the other agent. Facilitation consists of two main activities: the identification of the other agent's activity and helping the other agent, which would be realised by a task sharing.

So interferences would be an entry point to study the cooperation, but the goal is to find the sources of interferences. One way is to identify the result of the activity of each agent, i.e. the know-how and its interaction with the process what is called the *internal current representation* of the situation (ROI, French acronym for “Représentation Occurente Interne”). It is in the mind of the human agent, and recorded somewhere in the memory of the artificial agent. The ROI is made up of some different attributes: information (stemming from activities of information elaboration); problems (stemming from activities of diagnosis); strategies (stemming from activities of schematic decision making); solutions (stemming from activities of precise decision making); commands (stemming from activities of implementation of solutions) [1]. Interferences can appear after comparison between the ROI of each operator during direct exchanges, or between the ROI of one agent and the supposed ROI (ROI’) of the other agent (cf. **Fig. 1.** Cooperation between two agents). The internal common frame of reference (RCI, French acronym for “Représentation Commune Interne”) is made of exchanges triggered by these interferences. Three forms of interference management may be used by the agent: *negotiation*, *acceptance*, *imposition* [1]. These forms imply, for human agent, cognitive and communication costs which are different. The *negotiation* aims at reducing the differences between both ROI by modifying one of them, on the basis of explanations. The *acceptance* is the update of her/his/its ROI from the interpretation of the ROI of the other agent (ROI’). This acceptance is chosen when the cost of a negotiation is too important or when an agent wants to facilitate the activities of the other. The *imposition* corresponds to the opposite of the acceptance.

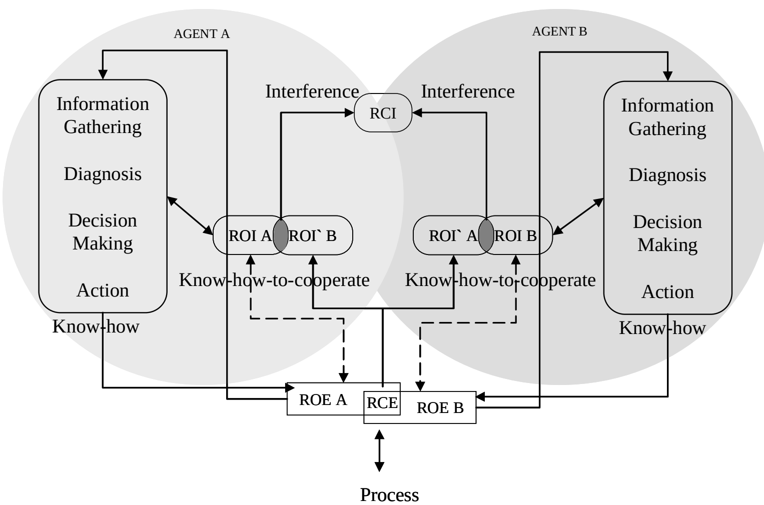


Fig. 1. Cooperation between two agents

One way to support exchanges, and so cooperation, is to implement the ROI on a machine, so called the *external common frame of reference* (RCE, French acronym for “Representation Commune Externe”). The Common Work Space (CWS) is based

on the RCE but more than gathering information, one goal of the CWS is also to make the know-how of each agent common [1].

As illustrated on the above drawing, the RCE can support the cooperation when an action has to be decided at a precise time, i.e. in the short-term. But if process plan is integrated on the RCE, cooperative activity corresponding to the long-term would be supported too. In fact, cooperative activities can be organized into three levels [5], at the same time implying an increase in the abstraction level and an enlargement of the time span:

- *Cooperation in planning* consists in elaborating or/and maintaining the RCI. It concerns common goals, common plans, role allocation, action monitoring and evaluation and, common representations of the environment. So, agents exchange information, problems, strategies, solutions and commands for sharing their/its own ROI.
- *Cooperation in action* groups together activities that have direct implications in the short-term and that rest on a local analysis of the process. This level includes local interference creation, detection, and resolution. It also integrates the anticipation of interference by identifying the goals of the other agents in the short-term. Cooperation in action consists also to actualize RCI.
- *Meta-cooperation,* situated at a much higher abstraction level, allows agents to improve the cooperative activities described above by elaborating long-term constructs, such as a common code to communicate easily and shortly, compatible representations formats, and above all models of one self and of the other agents. As said before, these models allow agents to build an interpretation of the ROI of the other agent (ROI'). Nevertheless this interpretation can be false.

The notion of common work space is already proposed by several authors in various work domains. In the case of Air Traffic Control, Bentley, Rodden, Sawyer and Sommerville [6] present a *shared work space* which provides an adapted presentation of the air traffic to different users, on different machines, to make the use of shared entities easier. In this work, the shared work space is limited to the external environment and process information representation. Other authors enrich the work space by different elements relative to the activity of each agent, and by not only raw information but also elaborated ones. Decortis and Pavard [7] define the *shared cognitive environment* as a set of facts and hypothesis which are a subset of the cognitive universe of each agent. In the specific field of Knowledge Based Systems, Brezillon and Pomerol [8] pointed out the necessity for a user and the system *to share a contextual knowledge* i.e. “the sharing of the cognitive representations, an agreement about the contextual knowledge of problem solving, the ability to follow each other's reasoning, the ability to exchange explanation and the control of the interaction by the user”. Royer [9] specifies that a system is all the more cooperative since the cooperation implies more levels as perception, analysis, decision and action. Karsenty [10] underlines the relevance of a *shared problem representation* to support the process of explanation during which co-workers articulate their individual problem representations and thus, build a richer shared problem representation, rend a deeper analysis of the decision space, and thus reach a better decision. Sonnenwald and Pierce [11] deal with *Interwoven Situational Awareness* which is “the continuous extraction of environmental information, integration of this information with previous

knowledge to form a coherent mental picture in directing further perception and anticipating future events”. They propose an interwoven situational awareness that includes the individual, intra group and inter group situational awareness for the Control and Command team members, who need to collect, synthesize and disseminate information to create an understanding of the current battlefield situation and to anticipate future battlefield events. Jones and Jasek [12] integrate this idea into the building of an *Intelligent Support for Activity Management* (ISAM) which allows sharing goals, contextual information and allocation of functions between the agents. Gutwin and Greenberg [13] highlight the importance to integrate the activity of the other agents and deal with a *workspace awareness* which is “the up-to-the-moment understanding of another person's interaction with the shared workspace”. The shared workspace allows knowing *who* is working, *what* they are doing, *where* they are working, *when* various events happen and *how* those events occur.

In this context, the goal of our study is to determine the impact of different type of CWS, based on the ROI structure, and on the cooperative activities in the cockpit. The first format is a temporal one dedicated to the short-term, and the second format is a spatial one, dedicated to the long-term. Our hypothesis is that the integration of a richer RCE could improve cooperative activity, but without implying a too important cost. An agent has: — to update the CWS on the base of her/his/its ROI, — to update her/his/its ROI on the base of the CWS, — to manage interferences between strategic and tactical levels in order to build the RCI and CWS.

These elements are evaluated in the fighter aircraft domain.

3 Fighter Aircraft Experiments

This study deals with the cooperation between a Pilot and a Weapon System Officer (WSO) of a two-seater fighter aircraft carrying out an Air-To-Ground mission. An *a priori* and strict task sharing exists between both members of the crew: the pilot is responsible for the short-term management of the situation (piloting, firing) whereas the WSO manages middle and long-term processes. An important part of the WSO work consists in setting up the work of the partner. In our experimental device the pilot as well as the WSO can perform every task (dynamic task allocation is possible like in actual work but in a larger extent in order to have more possibilities to study cooperation). A mission is divided into three main stages: (i) going near to the target (entering the enemy territory and flying to the target, steering clear of enemy threats), (ii) handling one or two targets (reconnaissance or destruction mission), and then (iii) leaving the enemy territory.

In collaboration with Dassault-Aviation, a generic two-seater fighter aircraft simulator was built. It is made up of three workstations, one station by operator and a third one dedicated to the management of scenarios and experiments, and the Wizard of Oz simulation.

The operators’ interfaces (identical for each crew member) are made up of: first, a big picture screen (cf. Fig. 2. (a) Big picture screen (b) short-term assistance tool) putting together the four main displays of a military aircraft cockpit, the assistance tools and some device controls (e.g. automatic pilot) that can be activated by the way of a touch screen. Second, the simulator can be mainly

controlled by a Hand on Throttle and Stick device, and by voice-control realized by the Wizard of Oz.

The simulation of a data-link communication allows displaying messages coming from a simulated AirBorne Control, Command and Communication (AB3C) by drawing or writing pieces of information about new threats or new targets. This big picture defines an RCE of the situation and it is already a shared work space because both operators see the same information and perceive in the same way the action of each other. The assistance tools aim at enriching this external common frame of reference.

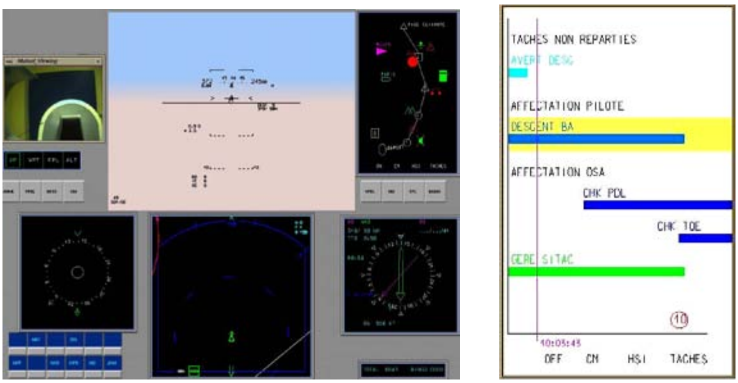


Fig. 2. (a) Big picture screen (b) short-term assistance tool

A long-term assistance tool (cf. Fig.2(a): upon part of the interface on the right) can be looked upon as an ROI of both agents. It is based on the screen they usually use in their today's cockpit. The content of this assistance is mainly made up by the crew during mission briefing. They can insert all information they want along a spatial representation of the flight plan. The information can be for example: — to verify if there are new enemy signalized by the partner on the CWS (information gathering), — to verify if the plane is well prepared for the next step of the mission, according to an automatic update of the plane configuration and of the mission progress presented on the CWS (diagnosis), — to decide to come home if the objective is unreachable (decision making), — to realize the video record of the objective to explore (action). During low workload periods of the mission, they can modify information displayed by voice command. So the crew builds the RCE of the situation in order to be ready if a crisis hits.

A short-term assistance tool, called a diary of major tasks (cf. Fig.2 (b)), can be looked upon as a common illustration of the state of the activity of each other for the next three minutes. As the long-term assistance tool, it is built during the mission briefing, and an automatic update of the progress state of the tasks displayed is realized. The progress state is coded by colours. When a task would not be completed at the proper time, the task is red, and an alert message is written in the main working window. This assistance tool provides more the actions and its states than information about the mission. In other words it is a more operational assistance tool with less

inference to do. It is also a CWS but it is focus on a short period of time, and so with the objective to support quick decision making and actions.

Another type of assistance tool consists in providing *the configuration of workstation of the partner*. One agent is able to know what the partner is able to see and/or to do. The aim of this assistance is to allow crew to make a new allocation of tasks if one control is broken, or to infer information the partner has. So, each operator can actualize the model of “technical availability” of the partner.

With the same idea, the *video picture of the face of the partner* is proposed. This assistance tool could help operators to know the state of the partner (overload, problem) by analyzing gestures, facial expressions, in addition to the voice which is already a good indicator. It could also allow the maintenance and the updating of the “capacity model” of the partner.

Operator can also *send short sentences to the partner*. A button on the stick allows the crew to use a voice control. A window allows to display the last 15 messages exchanged with a colour code to distinct the pilot's messages, from the WSO's messages, and from the AB3C's messages. The interest of this type of assistance is the persistence of the message, and avoids too important synchronization between both operators activity. One operator can read a message when she/he is available. It is a short-term assistance tool that allows anticipating the information to transmit. It could also be looked upon as an assistance tool to manage interferences, and to manage the lack of synchronization due to lack of communication (silence requested, breakdown).

Twelve crews of French Air Force of the military base 116 accepted to test these cooperative supports. The crews performed three missions in three conditions: the first one without assistance tool is the reference one, both assisted conditions provide operators with the exchanges of written messages, the configuration of the partner’s workstation and the video picture. But in the short-term assistance tool condition the diaries of major tasks is proposed, and in the long-term assistance tool condition it is replaced by the spatial representation of the mission. The scenarios were different but required similar tasks. The duration of a scenario lasted about 30 minutes. Order, rank, scenarios and conditions were counterbalanced using a greco-latin square experimental design. The spontaneous verbal reports and the video output of both workstations were collected. In addition, we used a method of self-observation reports after task execution. They had to answer to different questionnaires aiming at: characterizing the crews, obtaining information about the cooperation in the cockpit, evaluating the flight simulator, the scenarios and of course the assistance tools.

The results stem from the analysis of the answers to the questionnaires, the analysis of objective data about cooperative activity during mission completion, and assistance tool use. All the audio-recordings during the mission were also transcribed in order to provide verbal protocol and then coded.

For more than half of the operators, the *short-term assistance tool* favours the cooperation by aiding the mutual control. It seems to be useful to realize the task. The use and the presentation of information were fairly easy but the management of the tasks seems to ask too many time and too many resources to be achieved by the crew during the mission. For them, the updating of this assistance tool must be automated. It is used as a checklist to follow to achieve the mission in order not to forget any information gathering or diagnosis or action fulfilment. WSO ask to have this type of assistance tool for their medium and/or long term activities.

For half of the operators, the *long-term assistance tool* favours the cooperation. The experimental procedure implies to use it during the preparation of the mission. For nearly all the operators, it is used to perform the mission, and the presentation of the information appears to be satisfactory. The use of the display seems to be very easy. However the crews do not modify it during the mission. This tool gives the same information that the notes taken by the crew during the preparation of the mission on the paper map of the flight plan. This assistance tool spares one operator to ask several information to the partner. For the most part, the crews ask to group the short-term and long-term assistance tools in an existing visualization, the SITAC (French acronym for the Tactical Situation). It would be a CWS based on their actual display.

For half the operators who answer (only 16 out of 24 do), the *configuration of the partner's workstation assistance* favours the cooperation. But they indicate that it is not useful to realize the mission or they do not use it. However, some operators react to the word "EMPTY" by asking to the partner if he has a problem with his workstation and if he needs help to supply the lack of information.

All the crews think that the *video picture assistance* is not useful to realize the mission and it does not favour the cooperation. Only one WSO indicates that he used it when he lost his instruments, and with this video display he knew when it was possible to ask information to his partner: it was a way to evaluate his workload.

All the operators who answer (8 out of 24), the *exchanges of written messages assistance* does not favour cooperation. They also indicate that it is not useful to realize the mission or they do not use it. The experienced mission is not favourable to this type of exchanges. Verbal message is always dominating written message. However this means of communication would be essential when the radio exchanges are forbidden. It would also be a necessity if the partner is not human but artificial as a plane without pilot (UCAV).

The comparison between the three experimental conditions according to the coding of the cooperative activities shows that there is no deep modification of the cooperative activities of the crew. It could be because of the lack of training of the crew with the assistance tools, and the high cost implied by their use. Only WSO uses it probably because of his strategic role, contrary to pilot who he is not continuously involved in the short term management of the mission. As a consequence, the building of the RCI becomes more asymmetric and favours the WSO point of view.

The amount of the information gathering is also depending on the experience of the crew. The more experience the crews have, the more they use the assistance tools. This result probably comes from the fact that a more experienced crew has a better management of the mission and therefore has more time to read the information presented by the assistance tools. Low experienced crews use all their resources to try to realize the mission, because it is difficult for them and they have not sufficient training with the assistance tools to be helped by them. They seem to only take information they need to perform the current action. The WSO and pilot use the information from the assistance tools for their individual activities as well as for their cooperative activities.

4 Discussion and Conclusion

The main assistance tools used are the short-term assistance tools with the diary of tasks, and the long-term assistance tool; both define a CWS but with a different focus. All information must be integrated into an existing cockpit display; the most interesting display is the SITAC (for tactical situation). It can combine information provided by short-term and long-term assistance tools. But in the military domain operators are used to work with high technologies, and they usually have trust in them. Technologies improve their capabilities and/or abilities. They are perhaps more inclined to accept new assistance tool. A brief comparison with other domains underlines some differences which lead us to the next remark.

The idea would be to try to define the cooperation according to two dimensions, a “technical” one, and an “affective” one. The terms used are perhaps not the better, but the idea is that the technical aspects of cooperation would be based on the technical feasibility of the cooperation constrained by the definition of the tasks of each agent implied in the cooperation, and the organization of these tasks in order to allow an efficient task allocation. The affective aspect of the cooperation would be based on more human features like the willingness to cooperate, the trust in the cooperation [14] even if this notion can be also in interaction with technical aspects; it would be also the competition when an agent wants to do better than the other, not with the goal of attempting a better performance which is a technical aspect, but for her/his own satisfaction. These both notions of technical and affective aspects of cooperation can be connected to Hoc definition mentioned in the first part of the paper where interference management and facilitation were distinguished. The first one allows a detection of incompatibilities between internal data managed by each agent, but this detection is possible because agents have the common goal to facilitate the objective of the other, notion that is nearest of the affective aspect.

The interest in the distinction is to detect if the unused of an assistance tool is due to lacks in the technical aspects: the assistance tool provides bad solutions, or useless because there’s no need, or there’s no good synchronization with human agent activity. But an unused can also be due to lacks in the affective aspects: the social environment is sometime unfavourable because human agent could be replaced by artificial one, or they have no trust in the solution provided by the assistance tool, or in their interaction with it. Nevertheless, human agent can need this help.

The present experiment allowed having a fine coding of the individual and cooperative activity. All the steps of the analysis are not presented there and more details will be found in [15], but from these data analysis one hypothesis could be extracted: the cooperative activity could be based on the same type of activity as the process control activity. The activity of the agent with whom the cooperation is necessary could be considered as a process more or less predictable interacting with the process to supervise and control. If we apply the four steps of the know-how to the control of cooperative activity we can imagine that information gathering is the goal of an agent to gather information about an other agent directly or using the CWS, diagnosis leads agent to compare both ROI, decision making of the cooperative activity is to decide if exchanges are necessary and how (by negotiation, imposition, acceptance), and action is the application of the selected decision by interacting with the other agent directly or by actualizing the CWS.

Acknowledgments. We thank the Dassault-Aviation for their collaboration. The experimental field is the Air Force, especially Air Base 116. We thank 2/4, 1/4 Air Force squadrons for their welcome and their cooperation. We thank also the IMASSA-CERMA staff for their help, J.-M. Hoc and R. Amalberti for their advises during this study.

References

1. Lemoine-Pacaux, M.P., Debernard, S.: Common Work Space to support the Human-Machine Cooperation in Air Traffic Control. Control Engineering Practice, A Journal of IFAC 10, 571–576 (2001)
2. Rasmussen, J.: Skills, rules and knowledge; signals, signs, and symbols, and other distinctions in human performance models, IEEE Transactions on systems, man and cybernetics, SMC13, 3 (May-June 1983)
3. Millot, P., Lemoine-Pacaux, M.P.: An attempt for generic concepts toward Human Machine Cooperation, IEEE SMC, USA, California, San Diego (October 11-14, 1998)
4. Castelfranchi, C.: Modelling social action for AI agents. Artificial Intelligence 103, 157–182 (1998)
5. Hoc, J.-M.: From human-machine interaction to human-machine cooperation. Ergonomics 43(7), 833–843 (2000)
6. Bentley, R., Rodden, T., Sawyer, P., Sommerville, I.: An architecture for tailoring cooperative multi-user display. In: Proceedings of CSCW’92 (November 1992)
7. Decortis, F., Pavard, B.: Communication et coopération: de la théorie des actes de langage à l’approche ethnométhodologique. In: Pavard, B. (ed.) Systèmes Coopératifs de la modélisation à la conception. Toulouse, France (Octarès 1994)
8. Brézillon, P., Pomerol, J.-Ch.: Contextual knowledge sharing and cooperation. Le Travail Humain, tome 62(3), 223–246 (1999)
9. Royer, V.: Partage de croyances: condition nécessaire pour un système coopératif. In: Pavard, B. (ed.) Systèmes coopératifs de la modélisation à la conception, pp. 253–270, 129 Toulouse (Octarés 1994)
10. Karsenty, L.: Cooperative work: the role of explanation in creating a shared problem representation. Le Travail Humain, tome 63(4), 289–309 (2000)
11. Sonnenwald, D.H., Pierce, L.G.: Information behavior in dynamic group contexts: interwoven situational awareness, dense social networks and contested collaboration in command and control. Information Processing and Management 36, 461–479 (2000)
12. Jones, P., Jasek, C.: Intelligent Support for Activity Management (ISAM): An Architecture to Support Distributed Supervisory Control. IEEE Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 27(3) (May 1997)
13. Gutwin, C., Greenberg, S.: A Framework of Awareness for Small Groups in Shared-Workspace Groupware. Technical Report 99-1, Department of Computer Science, University of Saskatchewan, Canada (1999)
14. Rajaonah, B., Anceaux, F., Vienne, F.: Study of driver trust during cooperation with Adaptive Cruise Control. Le Travail Humain 69, 99–127 (2006)
15. Hoc, J.M., Lemoine, M.P.: Cognitive evaluation of human-human and human-machine cooperation modes in Air Traffic Control. The. International Journal of Aviation Psychology 8(1), 1–32 (1998)

Culture and Communication in the Philippine Aviation Industry

Rosemary Seva[1], Alma Maria Jennifer Gutierrez[1], and Henry Been Lirn Duh[2]

[1] De La Salle University - Manila, Philippines
[2] Nanyang Technological University, Singapore
sevar@dlsu.edu.ph, gutierreza@dlsu.edu.ph, mblduh@ntu.edu.sg

Abstract. This study aims to characterize the communication patterns of Filipino pilots in the cockpit and assess its adherence to CRM principles of assertiveness, openness, conflict resolution, and communication effectiveness. A survey was administered to 88 participants who are mostly captains and first officers. Filipino captains and first officers generally have the same attitude that corresponds to the principles learned in CRM training on communication and coordination. However, a detailed comparison of the groups' responses showed significant differences in terms of assertiveness, openness, and communication effectiveness. Captains are more articulate in expressing concerns than first officers. However, in terms dealing with crewmembers, first officers are more cautious in phrasing things.

1 Introduction

1.1 Background of the Philippine Aviation Industry

The Philippine aviation industry had been around for more than 50 years and was the first one in Asia. The flag carrier Philippine Airlines (PAL) was the pioneer in the industry and continues to be the strongest player up to the present time. PAL had been the only company that served domestic routes for many years until the airline industry was deregulated in 1995 paving the entry of other airline companies such as Air Philippines, Cebu Pacific, Southeast Asian Airline (SEAIR), and Asian Spirit (Deparment of Labor and Employment, 2006).

The operation of the air transport industry is subject to many regulations like other airline companies in the world. In the Philippines, the Air Transportation Office (ATO) implements relevant policies in civil aviation and oversees the implementation of safety procedures (Department of Transportation and Communication, 2003). Aviation companies strictly follow safety standards set by international organizations such as the Federal Aviation Administration (FAA) of the US and the European Aviation Safety Agency (EASA) of the European Union (Deparment of Labor and Employment, 2006).

1.2 Crew Resource Management and Communication

Like their international counterparts, aviation companies in the Philippines incorporate training programs to enhance pilot and crewmembers' safety awareness and

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 806–813, 2007.

behavior. Crew resource management (CRM) is one such program that was introduced in the early 1980's to prevent pilot errors (Fisher & Phillips, 2000; Helmreich & Foushee, 1993). The program was conceived due to the fact that majority of aviation accidents can be traced to human error brought about by lack of coordination among team members in the cockpit (Fisher & Phillips, 2000; Salas et al., 1999). Helmreich, Wilhelm, Klinect, & Merritt (2001) asserted that CRM training is an effective organizational strategy that may be used to minimize risk and manage human error.

Initial CRM training is comprised of several core elements, namely; human error and reliability, safety culture, organizational factors, stress management, information acquisition, situation awareness, workload management, decision making, communication and coordination, leadership and team behavior (Droog, 2004). The training emphasizes the need for improving team skills as a means to achieve operational effectiveness and safe service (Helmreich & Merritt, 2000). A study conducted by Fisher & Phillips (2000) showed that CRM training increases safety awareness and provides a safer operating environment.

Communication and coordination is an important aspect of CRM training. Training modules include effective communication skills, conflict resolution, assertiveness, communication styles, and relevant communication models that apply to pilots and cabin crew. These topics are taken up in order to mitigate the threats brought about by poor communication among crewmembers. Chute & Wiener (1995) highlighted the role of poor communication and coordination as the cause of recent airline incidents and accidents. They cited reports obtained from the NASA Aviation Safety Reporting System (ASRS) where the antagonism between pilots and cabin crew resulted in a near miss situation.

Communication problems in the cockpit can also be caused by the prevalence of a strong national culture as discussed by Jing, Lu, Yong, & Wang (2002) among Taiwanese pilots. Authoritarianism in the cockpit was identified as one of the threats to global flight safety as manifested by the high correlation of this variable to accident rates (Jing, Lu, & Peng, 2001).

1.3 Culture and Aviation

Pilot behavior is influenced by three cultures, namely; national, professional, and organizational (Helmreich, Wilhelm, Klinect, & Merritt, 2001). National culture has the strongest influence among the three because of the length of exposure to this factor. Merritt & Helmreich (1996) explained that pilot's attitudes especially in terms of command and communication were significantly related to dimensions of culture defined by Hofstede (1980).

Hofstede (1980) enumerated four dimensions of culture, namely; power distance (PD), uncertainty avoidance (UAI), individualism (IDV), and masculinity (MAS). PD refers to a belief and acceptance of unequal power distribution (Johnston, 1993). People from higher PD cultures such as Filipinos, Taiwanese, and Indians are more willing to accept authority and are afraid to question superiors out of respect or fear of punishment. Helmreich & Merritt (1998) and Hofstede (1980) placed the Philippines

among nations with high PD although Helmreich & Merritt's (1998) study only included pilots.

1.4 The Filipino Culture

Filipinos are characterized as friendly, cooperative, and very hospitable (Jocano, 1999). They are easy to get along with because they believe in the value of conforming to society. They are not individualistic and delight in the company of friends and relatives. They try to maintain harmonious relationship with one another especially within the group. As such, it is very easy to work with them and this is one trait that foreign employers value in the workplace (Jocano, 1999).

Filipinos put high value on positions of authority (Andres, 1989). Studies conducted to characterize Filipino culture showed that they have high power distance (Helmreich & Merritt, 1998; Johnston, 1993). They respect the authority of bosses in the workplace and parents at home.

The communication style of Filipinos is ambiguous (Jocano, 1999). Real intentions are not expressed directly and vague words are commonly used to avoid confrontation. Indirectness, however, is not used to mislead but to show consideration to other people's feelings. Words are not enough to communicate in the Philippine context. Body language and shared information are critical areas that facilitate understanding within society.

This study aims to characterize the communication patterns of Filipino pilots in the cockpit and assess its adherence to CRM principles of openness, assertiveness, conflict resolution, and communication effectiveness. Since national culture is a dominant variable in shaping an individual's behavior, it is hypothesized that cultural influences prevail in the communication style in the cockpit.

2 Method

2.1 Participants

A total of 88 experienced Filipino aircrews from four airline companies in the Philippines participated in the survey. All the participants were male who were mostly captains (40%) and first officers (41%). The rest were managers, check pilots, or flight engineers. Majority of the sample (82%) flies domestic routes and the rest flies to international destinations. Seventy five percent of the participants were trained in pilot school and the rest were trained in the military.

Ninety seven percent of captains and 100% of first officers in the sample have undergone CRM training. Sixty four percent of them have undergone either a one- or two-day seminar on the topic.

In this study, only results obtained from captains (35) and first officers (44) were considered for the purpose of comparison.

2.2 Questionnaire

Data were gathered using the Human Factors and Safety Culture Survey developed by Lu et al. (2000). The questionnaire has three sections: CRM concept, CRM commitment,

and CRM skill. Each of the sections was subdivided into six subsections, namely; communication, situational awareness, teamwork, workload management, decision-making, and culture. The questionnaire consists of 88 items that pilots rated using five-point rating scales of agreement or frequency. As a whole, the questionnaire evaluated 58 variables related to CRM implementation such as stress understanding, company training, language skill problem, perceived safety record, etc. (Jing, Lu, Yong, & Wang, 2002) The detailed questionnaire can be found in Lu et al.(2000).

The questionnaire considered cultural factors such as culture barriers in training and power distance, however; only ten variables that relate to communication and culture were analyzed in this study.

The questions analyzed are summarized in **Table 1**. They were classified into four groups that correspond to CRM principles where pilots are trained in CRM.

Table 1. Communication Questions Analyzed

	Assertiveness
A1	If I perceive a problem with the flight, I will speak up, regardless of who may be affected
A2	Crewmembers should not question the captain, except when they threaten the safety of the flight
	Openness
O1	I am reluctant to express disagreement with my superiors
O2	I feel comfortable going to my manager's office to discuss problems or operational issues.
O3	I am encouraged by my supervisors to report any operational problems I may observe.
	Communication effectiveness
E1	When I am communicating to my superiors, I need to be careful how I approach him
E2	When I am speaking to my crewmembers, I need to be careful how I phrase things.
E3	My superiors are courteous and polite when they are talking to me
	Conflict resolution
C1	If I have a problem with the actions of my crewmembers, I will tell him so and attempt to resolve it
C2	When my crewmembers object to what I do, I am willing to discuss it with them

3 Results

Responses to the ten communication questions were grouped according to captains (C) and first officers (FO) to enable comparison. The mean and standard deviations for each of the ten questions are presented in **Table 2**.

Table 2. Descriptive Statistics

		Captain		First Officers	
	Assertiveness	Mean	SD	Mean	SD
A1	If I perceive a problem with the flight, I will speak up, regardless of who may be affected.	4.60	0.50	4.30	0.63
A2	Crewmembers should not question the captain, except when they threaten the safety of the flight.	3.60	1.22	3.50	1.41
	Openness				
O1	I am reluctant to express disagreement with my superiors.	2.69	1.08	2.90	0.92
O2	I feel comfortable going to my manager's office to discuss problems or operational issues.	4.00	0.94	3.43	1.13
O3	I am encouraged by my supervisors to report any operational problems I may observe.	4.35	0.69	4.05	0.94
	Communication effectiveness				
E1	When I am communicating to my superiors, I need to be careful how I approach him.	3.03	1.20	3.52	0.99
E2	When I am speaking to my crewmembers, I need to be careful how I phrase things.	2.91	1.15	3.64	0.98
E3	My superiors are courteous and polite when they are talking to me.	4.09	0.51	4.26	0.66
	Conflict resolution				
C1	If I have a problem with the actions of my crewmembers, I will tell him so and attempt to resolve it.	4.00	0.87	3.69	0.95
C2	When my crewmembers object to what I do, I am willing to discuss it with them.	4.14	0.85	4.07	0.93

As can be seen from the results, both groups think that there is a need to speak up when a problem is perceived during a flight. There is also a weak belief that the authority of the captain is absolute and should not be questioned as manifested by the mean rating of less than 3.75 for question A2.

Captains are less reluctant than first officers in expressing disagreement with superiors (O1). They also exhibited more openness in discussing and reporting operational problems (O2 and O3).

In terms of communication effectiveness, it is apparent that first officers are more careful in dealing with their superiors and crewmembers (E1 and E2). The general result, however, indicates that both groups are not particular with being cautious in the manner of communicating with superiors or subordinates. Captains show indifference

in these two issues as shown in mean ratings close to 3 (“neutral”). However, both groups felt that their superiors are polite and courteous in dealing with them (E3).

Captains have higher ratings in questions dealing with conflict resolution than first officers. They are more open to discuss disagreements and confronting crewmembers about problems in their actions. This kind of behavior is expected of captains, as they are the ones who are in charge of the flight and acts as the manager of the team. This also shows a positive transfer of CRM training principles in the area of communication and coordination. Although first officers showed a willingness to discuss disagreements with crewmembers, they are less willing to confront problems with crewmembers and resolve them.

One way Analysis of Variance (ANOVA) was conducted to identify significant differences between the two groups for each of the questions considered. Results of the ANOVA revealed significant mean difference for questions A1 ($p<0.05$), O2 ($p<0.05$), and E2 ($p<0.01$).

4 Discussion

The survey unearthed unique communication patterns among Filipino pilots, specifically captains and first officers. Both groups are assertive in communicating perceived problems despite of possible ill repercussions. It was also observed that they do not put high value on authority which is not typical of the Filipino culture of high power distance (Helmreich & Merritt, 1998). Such attitudes may be attributed to the influence of CRM teachings in the area of communication and coordination since almost all of them have attended the training very recently (less than three years). Thus, it is logical to assume that basic principles taught are still fresh in their mind and have influenced their way of thinking.

There was a significant difference observed in the attitudes of captains and first officers in A1. Captains are more articulate in addressing problems probably because they consider themselves the manager in the flight. Although the general results suggest that first officers are also willing to speak up when they perceive problems they are still less assertive compared to the captains because they still defer to their authority. The first officers show the value of “paggalang” or respect that is important in interpersonal relations in the Philippines (Jocano, 1999). Since Filipinos belong in a non-individualistic society, they value relationships with others especially their superiors in the company.

Captains are more comfortable in discussing problems with managers than first officers and the difference in their response is statistically significant. This may be attributed to the good organizational culture that prevails in aviation companies in the Philippines and the norms followed by captains in the industry. The survey showed that captains feel that their superiors are courteous and polite when talking to them and this attitude by the companies’ management may have influenced their good outlook towards communication. Since captains are also higher than first officers in the hierarchy of authority, it is expected that they are more comfortable in dealing with the company’s management because it is part of their job.

Both captains and first officers are not mindful of being cautious when communicating with superiors and subordinates although first officers are significantly more

guarded in their communication style. The general trend can be explained by typical Filipino behavior where "kapwa" or "being on equal terms with others" is an imperative (Jocano, 1999). Filipinos who practice the "kapwa" standard is positively evaluated in society. This standard keeps the hierarchy in society intact but insists on treating everybody on equal terms as a fellow human being or fellow worker.

Generally, captains and first officers have a positive attitude towards conflict resolution. They are willing to discuss objections raised by crewmembers. This is a typical Filipino attitude that tries to avoid conflict. In the Philippines, leaders are valued for their sobriety (Andres, 1989). They must resolve conflict without offending anyone. It usually requires charm and good personality. Another important aspect of good behavior in the Philippines is "pakikitungo" or being civil and relating to others the most appropriate way possible.

The significant differences in attitudes between captains and first officers in the survey are brought about by strong national culture that influences their behavior. Although principles of CRM have already been introduced to them it is still very difficult to deviate from the dictates of accepted social norms. However, general trends in response show adherence to CRM principles of communication and coordination.

5 Conclusion

Filipino captains and first officers generally have the same attitude that corresponds to the principles learned in CRM training on communication and coordination. However, a detailed comparison of the groups' responses showed significant differences in terms of assertiveness, openness, and communication effectiveness. Captains are more articulate in expressing concerns than first officers. However, in terms dealing with crewmembers, first officers are more cautious in phrasing things.

The Filipino culture is a very potent influence in the cockpit that is moderated by learning obtained from CRM. Even in an environment where professional culture is strong, the effect of national culture prevails in the area of communication and coordination.

References

1. Andres, T.: Management Filipino Style. Makati, Philippines: St. Paul (1989)
2. Chute, R.D., Wiener, E.L.: Cockpit-cabin communication: I. A tale of two cultures. The. International Journal of Aviation Psychology 5(3), 257–276 (1995)
3. Deparment of Labor and Employment, The air transport industry: think paper (2006) Retrieved May 23, 2006 from http://www.dole.gov.ph/nationalsummit/PS%20-%20Aviation%20Industry.pdf
4. Department of Transportation and Communication. Sectoral and attached agencies (2003). Retrieved May 22, 2006, from http://www.dotc.gov.ph/actioncenter/agency/airtransport.htm
5. Droog, A.: The current status of CRM training and its regulation in Europe. In: Goeters, K.-M. (ed.) Aviation Psychology: Practice and Research, pp. 221–230. Aldershot, UK: Ashgate (2004)

6. Fisher, J., Phillips, E.: Does crew resource management work? Air. Medical Journal 19(4), 137–139 (2000)
7. Helmreich, R.L., Foushee, H.C.: Why crew resource management? Empirical and theoretical bases of human factors training in aviation. In: Wiener, E.L., Helmreich, R.L., Kanki, B.G. (eds.) Cockpit Resource Management, pp. 3–45. Academic Press, San Diego (1993)
8. Helmreich, R.L., Merritt, A.C.: Culture at work: National, organizational, and professional influences. Aldershot, UK: Ashgate (1998)
9. Helmreich, R.L., Merritt, A.C.: Safety and error management: The role of crew resource management. In: Hayward, B. J., Helmreich, R. L., Merritt, A. C. (eds.) Aviation Resource Management, pp. 107–119. Aldershot, UK: Ashgate (2000)
10. Helmreich, R.L., Wilhelm, J.A., Klinect, J.R., Merritt, A.C.: Culture, error and crew resource management. In: Salas, E., Bowers, C.A., Helmreich, R.L., Wilhelm, J.A., Klinect, J.R., Merritt, A.C. (eds.) Improving teamwork in organizations, pp. 305–331. Erlbaum, Hillsdale (2001)
11. Hofstede, G.: Culture's consequences: International differences in work-related values. Sage, Beverly Hills (1980)
12. Jing, H.S., Lu, C.J., Peng, S.J.: Culture, authoritarianism and commercial aircraft accidents. Human Factors and Aerospace Safety 1(4), 341–359 (2001)
13. Jing, H.S., Lu, P.J., Yong, K., Wang, H.C.: The dragon in the cockpit: the faces of Chinese authoritarianism. Human Factors and Aerospace Safety 2(3), 257–275 (2002)
14. Jocano, F.L.: Working with Filipinos: A cross-cultural encounter. MPC Printers, Quezon City, Philippines (1999)
15. Johnston, N.: CRM: Cross-cultural perspectives. In: Wiener, E.L., Kanki, B.G., Helmreich, R.L. (eds.) Cockpit Resource Management, pp. 367–398. Academic Press, San Diego (1993)
16. Lu, P.J., Wang, S.E., Co, H.J., Jing, H.S., Yuan, S.F., Dai, M.: Cooperate Resource Management and Corporate Culture Research Report. Taiwan: Aviation Safety Council (2000)
17. Merritt, A.C., Helmreich, R.L.: Human factors on the flight deck: The influence of national culture. Journal of Cross-cultural Psychology 27(1), 5–24 (1996)
18. Salas, E., Prince, C., Bowers, C.A., Stout, R.J., Oser, R.L., Cannon-Bowers, J.A.: A methodology for enhancing crew resource management training. Human Factors 41(1), 161–172 (1999)

Future Trends in Flight Deck Equipment

Alison Starr[1] and Piet Hoogeboom[2]

[1] Smiths Aerospace, Bishops Cleeve, Cheltenham, UK
[2] National Aerospace Laboratory
Anthony Fokkerweg 2, 1059CM, Amsterdam

Abstract. This paper discusses work undertaken in the HILAS (**H**uman **I**ntegration into the **L**ifecycle of **A**viation **S**ystems) project which has analysed key drivers for change in aviation. The required technological developments and potential technologies that could be introduced to the flight deck to support these changes are elaborated upon, starting from a historical perspective to the development of flight deck technologies.

Keywords: Flight deck technology; trends in aviation; historically development of the flight deck.

1 Introduction

The world of aviation is changing and there are a number of external factors that drive these changes. These factors have an impact on the environment in which the aircraft operates and on the way in which the aircraft needs to be operated. These changes will require new technology to support them. This paper discusses some of these driving factors and how candidate technologies are being investigated (e.g. in the context of the EC project HILAS) that could support the future needs of the industry.

Within the HILAS project a strand of work is dedicated to flight deck technologies. It is primarily work from this that is discussed here. Within the project the technologies are used as case study material to test new tools and processes for future design. This paper however will discuss the trends identified by HILAS and outline some supportive technologies.

2 Driving Factors for Change

In order to identify and understand the technology changes being envisaged it is essential to understand the requirements in aviation and the world in general behind them. The driving factors for change and development across the industry are complex and interlinked. While there are many this paper will focus on the key drivers and discuss how these are linked to potential technology requirements on the flight deck.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 814–823, 2007.

The key drivers to be discussed are:

Economic – reducing cost
Congestion – too many aircraft, too little space
Environment – going green

These factors have been analysed from a number of different sources including material from within the HILAS partner organisation. An external and widely accepted reference that aligns with many of these factors is the ACARE Vision 2020 document [1].

2.1 Economic – Reducing Cost

Increasingly the population of the world is taking to the air and travelling for work and for pleasure. However, this travelling population is demanding and the market to service their needs is highly competitive. One of the main measures on which an airline is measured is cost, as reflected in the ticket price. Therefore one of the key drivers for the future is the need to reduce costs. A number of changes have taken place in recent years. Especially noteworthy is the introduction of low cost, no frills airlines such as Easy jet. Increasingly ticket price, and not quality of service, has become the primary reason for airline selection by the passenger. Thus all operators are looking for changes that will reduce cost.

The cost of the industry is made up in many parts from the cost of oil; cost of crew; the cost of the aircraft and of its maintenance. In order to reduce the cost of operating an aircraft all of these factors need to be considered. There are external pressures on many of these factors for example oil reserves are decreasing and it is possible that they will run out in the next 30 years. As the supply reduces the cost for this commodity will increase. Cost increases are expected to grow even be faster due to the current political focus on environment protection. Hence the need to reduce fuel consumption is a primary goal of the industry. From a flight deck perspective technologies which can support the refinement of the flight profile and path of the aircraft to reduce path length and to fly the ‘ideal’ path for the aircraft could form part of the solution. This could be supported by improved accuracy in weight and balance calculations. The technologies to achieve this might include new flight management functions and supportive technologies such as those found on *electronic flight bags.*

After cost of the fuel another significant cost is that of the people required to operate the aircraft. On the aircraft itself there is crew in the cabin and those on the flight deck. The number in the cabin is fixed by safety related regulations, and is especially related to the need to evacuate the passengers in emergency. However, on the flight deck there has over the decades, been a reduction in crew numbers, supported by technology change. Looking back in time the initial large passenger aircraft carried a crew that could have included an engineer, a radio operator, and navigator, in addition to two pilots. However, technology has slowly replaced these none flying roles leaving two pilots as the standard crew. Could this trend be continued to one or even no crew (unmanned aircraft)?

There is a trend on larger ultra long haul aircraft already for some phases of flight to be handled by only one crew member; or a situation were a second crew works in the cruise phase only, while the ‘taking off and landing crew’ rest. Without the need to take off or land the skill level of the ‘cruise’ pilots could possibly be reduced. Allowing single crews to operate in these phases of flight would have an impact on cost of operations, through reduction in personnel cost and in the cost of training. The next move could be to a single crew aircraft with only systems for back up. This would further reduce training and personnel costs. It could also potentially reduce the cost of the aircraft as there would be a decrease in the number of flight deck systems required. However these savings need to be balanced against the cost of new technology and training required for support the lone crew member.

This goal may not be a million miles away as fighter aircraft currently operate with single crew, and small business aircraft can also be operated by single crew members. However the carrying of passengers and the extended safety constraints make the environment very different. New technology that makes the workload manageable by a single crew member at all times, and that can support all crew tasks, and situations (including system failures) is required. This could include *information providing systems, potentially increased automation and support systems, and different displays and controls that can present the output of the new systems application***s.**

2.2 Congestion – Too Many Aircraft, Too Little Space

Airspace is affected by a number of factors. The number of aircraft operating clearly has a significant impact. There has been an almost continuous growth in air traffic over the decades. Historically aircraft where free to flying the open airspace, but it became apparent that this free use of space was far from safe, especially as airports became structured places that a large number of aircraft need to land at. This leads to the introduction of aerospace in which aircraft are controlled from the ground. Now this controlled space is filling especially on the approaches and departure routes to the airports. Runway capacity is at maximum at many main airports and air traffic controllers are working at full capacity. Other factors such as weather, restricted over flying and over fly charges, all add to the congestion equation.

While currently the system as a whole is still working, there are predictions that the demand on the system will continue to increase. Figures for past and predicted revenue passenger kilometres (see Figure 1) show that there will be an increasing demand. While there have been occasional decreases in demand from passengers, these can be related to specific global events and have a short-lived impact on the demand. For example there was a drop in demand after 2001 due to the attack on the twin towers in New York in 2001, and in early 2003 an outbreak of Severe Acute Respiratory Syndrome (SARS). But recovery was predicted as almost immediate, and from that point rate of increase continued to expand.

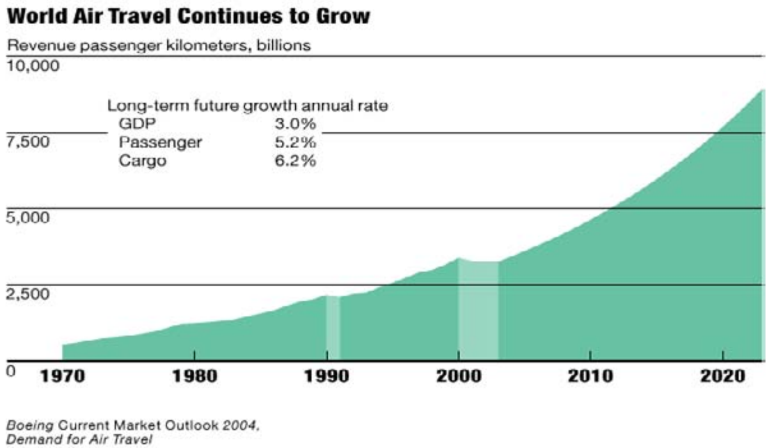


Fig. 1. World air travel developments (http://www.boeing.com/commercial/cmo/2-1.html) [2]

In fact detailed figures (Table 1) show that on all major measures recovery continues. The figures for Schiphol Airport Amsterdam between 2003 and 2004 reveal an increase for passenger; cargo; and for aircraft movements at 6.5%, 8.8% and 3.3% respectively. These levels of demand are already higher than pre 2001.

Table 1. Key Figures for 4 major European hubs for 2003 and 2004 [3]

Airport	PX 2003 (million)	PX 2004 (million)	Cargo 2003 (1000 ton)	Cargo 2004 (1000 ton)	A/C movements 2003 (1000)	A/C movements 2004 (1000)
London	112	119	1647	1721	863	889
Paris	70	75	1589	1740	709	735
Frankfurt	48	51	1527	1724	445	463
Amsterdam	40	42,5	1307	1421	393	402

This demand for more space can either be met by enabling more movements or by increasing the size of the aircraft themselves. Both of these options are being actively pursued by the airlines and the business as a whole. Opening up higher altitudes would be one way of increasing capacity, another currently receiving a great deal of research attention would be to change the air traffic management system to one in which the aircraft has more freedom to move and has responsibility for its own separation. Such a change would have significant impact on the roles of the pilots, and on the flight deck would require information about the future path of other

aircraft, for control and management of information relating to the flight path changes etc. This would require *new display and control technologies.*

Another area where new technology could aid the throughput of aircraft at airports, and thus speed up the system and increase capacity over all is by support taxiing. Currently taxiing is slow and can lead to slowing down overall movements at airports. This is particularly the case at complex airports or where weather is poor. At NASA AMES [4] research has been carried out on providing the crew with 'visualisations' of taxi information. Their results show that this not only helps prevent runway incursions but reduces taxiing time by about 16%. Therefore new technology that offers a *taxi guidance system* could enhance airport usage by allowing more aircraft movements. Taxi guidance systems might be as simple as a follow me car or lights, or a moving map, or head up display possibly with added information.

2.3 Environmental – Going Green

The environment and sustainability are key topics for the future of the industry. There is considerable demand for a reduction in both emissions and noise, around airports and beyond. New technology has a key role to play in achieving these aims. In addition to work on the engines, the control of noise impact could be improved by changing route of the aircraft to avoid the most built-up areas, and using procedures designed to minimise the noise heard on the ground. Aircraft emissions including gases such as CO_2 and NO_2 are an issue for the environment. Techniques and technologies that might alleviate emission include advanced engine design, forcing aircraft to be come more effective through varying landing fees at an airport based on fuel consumption, and the emissions of the aircraft itself. Alternatively flying a route at the optimal altitude and path for the current aircraft configuration (weight, speed, destination etc.), considering both energy and noise is one significant method of achieving goals associated with the environment. This means new technology on the flight that can calculate these *optimised routes, formats to display such information and display and control technology* that is capable of presenting this to the crew, including improvements to the flight management system and autopilots.

3 Technology Changes

These requirements and challenges to support the changing world of aviation are set in the context of technology that is rapidly changing. One in which historically aerospace was a technology leader, a discoverer of the next global step in many areas. However, today the commercial world, especially of computing and communications technology, moves at a rapid pace. In the world of commercial electronics the size of the market, i.e. you and me, every home, is vast. Therefore the profit to be gained by improvements, developments and the introduction of the next exciting new technology is significant. This produces a timetable for change that drives technology development that can not be matched by aerospace. However these technologies can be adapted and further developed to service the needs of the aerospace industry. We have become adaptors and users of technology in many areas, in particular on the

flight deck. Taking technology in the form of new displays (e.g. Liquid Crystal Displays) and control concepts (such as touch screens and speech inputs), and adding tailored applications we can develop systems that give our business the cutting edge solutions required by the aircraft manufactures and airlines.

The flight deck from which the pilots operates the aircraft developed from the original open cockpits operated through a pair of mark-one human eyeballs. A flight deck where the only instrumentation that might seen were for the engineers. For example the only instruments fitted to the Wright Flyer in 1903 were used to record the performance of the aircraft so that the engineers could use the data collected to make comparisons with their theoretical predictions [5].

Pilots soon began to realise that flight information would be of value to them in supporting their tasks. The first ‘instrument’ was probably a piece of string tied to the airframe, which could be monitored by the pilot as an indication for side-slip. The first panel instrument was probably a engine performance indicator, displaying RPM, see Figure 2, adapted from the engineers instrument and given to the pilot as engine problems caused many of the early accidents. Soon other instruments, like speed and attitude were added to help the pilot keep the aircraft airborne. The ‘clock work’ technology available in the early days meant that many instruments were adaptations of technology currently in use, like barometers and watches. The first altimeters were aneroid barometers with new faces; however this meant that as your aircraft climbed the dial moved anti-clockwise (downwards!)[5]. It was soon recognised that for any display system to be useful, without causing the human users to make errors, it must provide information which rapidly assessable, with a high probability that the assimilation of that information is correct.

The next step in information provision on the flight deck saw the development of instruments to support flying without visual reference to the external environment. In 1929, when James Doolittle used a gyro horizon, directional gyro and sensitive altimeter to fly ‘blind’[5]. These instruments, added to the airspeed indicator and the turn and slip indicator, formed the centre piece of the flight deck that we still see today.

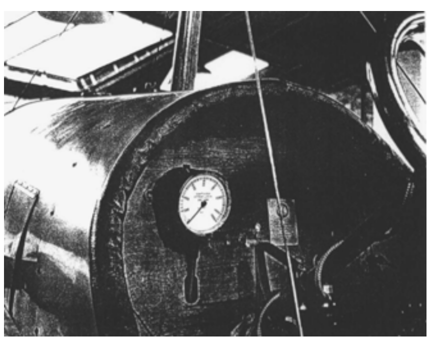


Fig. 2. Blackburn Monoplane with Single Engine Indicator (circa 1912)

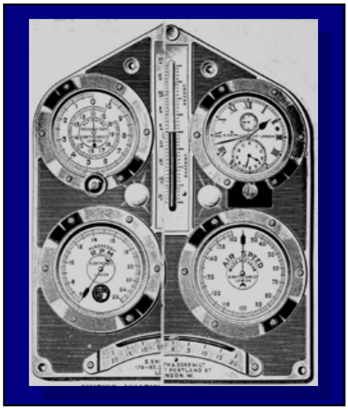


Fig. 3. Early Smiths Instrument Panel

Since then a large number of innovations and development have been seen. The drivers for change are now linked to increasing the profit of the airlines. The drivers for change discussed above lead to some potential technology needs for the future. These include the need for **new applications** to support optimized routing, information to support the tasks of a single crew member, and control and output technologies in the form of **electronic flight bags and new display and control technologies.** The rest of this paper briefly discusses some of the candidate technologies being investigated, in HILAS, to fill these needs.

3.1 Future Display Technologies

Over the passed decades electro-mechanical displays have been reduced by multi function ‘glass’ displays. First in the form of Cathode Ray Tubes (CRT) followed by the Liquid Crystal Displays (LCD). The information once presented on individual bespoke instruments is now integrated into a single display. While the display technologies of the future (such as organic light-emitting diode (O-LEDs)) may offer benefits in terms of reduced unit cost, lower maintenance and operating costs, etc., it is the information, and the nature of the information, they provide to the crew and consequent the way in which that information allows them (or a single pilot) to operate the aircraft that is likely to have the most significant impact in the quest to meet further needs.

As display technology has developed we are now able to produce even larger display surfaces on which to present the information to the crew. One of the options for future flight decks is to introduce larger displays surfaces with different formats that contain more and different information. For example information on the location of other aircraft for assuring self separation. Other display possibilities are the introduction of dual layer displays that enable new forms of 3D synthetic vision applications. This means an additional information dimension that can be presented to

the crew. In this display this is used to show distance along the flight-path[1]. This may well enhance the ability of the pilot to focus on the salient elements of information at the required time and to more intuitively process the information within the format. These features are being investigated as part of the HILAS research programme.

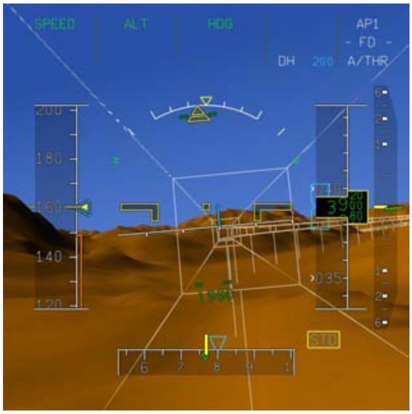


Fig. 4. 3D Synthetic Vision display with a Tunnel-in-the-Sky guidance1

Another display technology set to make potentially make an increasing mark on the civil flight deck is the Head-Up Display (HUD). These have been widely used in military aviation but have also found a place in civil aircraft. This has to date been in the business-jet sector where the aircraft owners are less price-sensitive and also in airlines where the use of the HUD enables them to operate in weather conditions (especially low visibility) at airports that are not equipped with category 3 landing systems. Current civil HUDs are used for critical information during landing when the pilot is operating head up and eyes out of the cockpit in the real world (see figure 5). The expansion of the role of the HUD is currently being researched and could extend its use to taxiing and ground handling.

Fig. 5. Civil Aviation Head Up Display

[1] This is a 2 D picture of the format seen on the dual layer display in the HILAS programme.

A display technology that has its value, in supporting the drivers for future aviation, entirely in the information and functionality it provides, is the Electronic Flight Bag (EFB). The EFB comes in many forms, for the cabin as well as the flight deck. The EFB is multi function it can display many different types and formats of information. Its application to provide data on, and to support the calculation of flight required functions such as performance data, fuel calculations, etc. is important in optimising the flight path of the aircraft. They can also be used as data/information storage devices holding information on for example:

- Performance calculations
- Weight & balance calculations
- Maintenance
- Charts etc.

Therefore EFBs may also save flight deck weight, and thus reduce fuel required, by replacing some or all of the hard copy material that pilots typically carry in their flight bags or have stored on the flight deck. Since EFBs also allow faster access to the immediately required information (e.g. checklists and aerodrome information), some safety benefits may be gained. Lastly, due to use of electronic means to transfer information to the EFBs, less paper is consumed leading to a reduction in wasted material, whilst ensuring a less error-prone information update process.

3.2 Control Technologies

On the flight deck the technologies that control the input of commands and data/information are also changing due to technology developments. These charges aim to better support the operator, reduce workload, increase situational awareness and have a positive impact on the operation of the aircraft, potential enabling single crew options. With control, more than with display, it is the way in the technology supports the human tasking that brings about the benefits. One example is the use of speech systems, particularly speech input. This technology allows the pilot to control systems, to input data and request functions on the flight while working eyes out, or while their visual attention is focused on another system or display. Another enabling control technology could be touch input. The change from hard fix function buttons and controls found on traditional flight deck to a flexible interface matched to the current task of the crew is enabled by the introduction of touch control, as it can be formatted to meet those current needs. One of the historic draw backs of traditional touch systems has been the lack of physical feedback to the pilot. However, new touch technologies are emerging that may overcome this and could give the pilot the long desired feedback in the near future. If this is achievable the door is open to significant changes in the way in which information is presented and controlled on the flight deck. Information could be truly provided to match the current tasking and needs of the crew, hopefully significantly reducing crew workload and increasing situation awareness. This form of information control and presentation is currently under research within HILAS and known as task oriented design. Research so far indicates that this approach does facilitate better situation awareness and understanding. The operator can complete tasks in a more naturally flow and consequently they are less likely to omit actions etc.

This should lead to improvements in safety, operational efficiency and potential reduce human errors.

4 Conclusions

In conclusion it should be said that there are many drivers for change in the aviation industry today. The drivers are interrelated and complex. Fortunately there are a raft of technologies that may in the future also offer solutions. This paper has focused on key drivers and resulting requirements, and outlined some of the flight deck related technology changes which have in the past, and should in the future enable there attainment.

Acknowledgments. This paper has been prepared with information from the HILAS project Strand 3 team.

References

1. ACARE concepts & SRA (strategic Research Agenda)
2. Boeing Statistical Data, http://www.boeing.com/commercial/cmo/2-1.html
3. Schiphol Airport data, http://www.schiphol.nl/media/portal/schiphol_group/pdf/pdf_files/fin_jv_2004_nl.pdf
4. Hooey, B.L., Foyle, D.C., Andre, A.D.: The design of aircraft cockpit displays for low-visibility taxi operations. In: Gale, A.G. (ed.) Vision in Vehicles IX, Elsevier, Holland (2001)
5. Chorley, R.A.: Trends in Flight Deck Displays. Smiths Aerospace, TSP 2674 (1977)

Evaluation of Process Tracing Technique to Assess Pilot Situation Awareness in Air Combat Missions

Ketut Sulistyawati, Yoon Ping Chui, Yeow Min Tham, and Yeow Koon Wee

Centre for Human Factors and Ergonomics
School of Mechanical and Aerospace Engineering
Nanyang Technological University
50 Nanyang Avenue, Singapore 639798
sulis@pmail.ntu.edu.sg, mypchui@ntu.edu.sg,
thamym@singnet.com.sg, weey0001@ntu.edu.sg

Abstract. This paper evaluates a process tracing technique developed to assess pilot Situation Awareness (SA) in air combat missions. Using the tool, the assessment of pilot SA and translation of pilot behavior to an observer assessment form is done through observation and analysis by Subject Matter Experts. The tool is found to impose minimal interruption to the task performance, which is desired from an assessment tool that is to be applied in a highly dynamic environment such as air combat. It is also found to be advantageous in providing comprehensive and detailed information of the dynamic changes in pilot SA and the situation assessment process, as well as the quality of SA acquired. Some problems with the tool are identified and modifications to minimize them are proposed.

Keywords: situation awareness, aviation, air combat, method, process tracing.

1 Introduction

Developing and maintaining a good Situation Awareness (SA) is highly critical in the aviation domain. Problems with SA have been identified to be a major causal factor in both military and commercial aviation mishaps [1]. SA is defined as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [2]. Specifically for air combat environment, US Air Force operationally defined SA as "a pilot's continuous perception of self and aircraft in relation to the dynamic environment of flight, threats, and mission and the ability to forecast and then execute tasks based on that perception" [3]. In air combat, the ability to obtain a higher SA would determine winning and losing in the battle. Enhancement of SA is therefore of high interest. Various interventions, such as personnel selection, system design, and training, are done as attempts to result in higher pilot SA. A valid assessment of pilot SA is thus needed to evaluate the effectiveness of these interventions.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 824–833, 2007.

Prior to the deployment of any assessment tool, an understanding of the pilot's task in air combat mission is necessary. This provides several implications on the requirements that need to be fulfilled by the SA assessment tool, namely:

1. Practical, with minimum level of intrusion to task performance. The air combat environment is highly dynamic. The fighter pilots need to obtain and process vast amount of information to develop and maintain the awareness of many complex and dynamic events that occur simultaneously in flight. Using an assessment tool that potentially interfere with this process, such as stopping (freezing) the simulations several times during task performance, may alter the task nature.
2. Capture the dynamic changes of SA. The faster the environment changes, the faster one's SA has to be updated. Developing and maintaining a pilot's SA is synonymous with being involved in a three-dimensional spatial relationship that is further complicated by the fourth dimension of time compression. A pilot may have perfect SA at one time and completely losing it at another time if he does not update his SA at the same pace as the system's dynamic. Therefore, capturing the changes in pilot SA would provide useful information regarding his ability to develop, maintain, and recover from lost of SA.
3. Assess the achievement of different stages of pilot SA. Air combat situation assessment involves the loop of "*search - detect – perceive – interpret - project*". The pilot needs to scan and search for contacts and information from various sources of information, such as radar, visual lookout, tactical data link displays, and voice communication with ground and airborne controller and friendly aircrafts. This information is then analyzed to obtain further information of enemy status in relation to own or friendly aircraft in terms of relative status (e.g., relative speed, distance, closure rate, and threat level) and predict enemy's intentions and actions, which may affect the pilot's decision of actions. Capturing the stages of SA achieved by the pilot can determine the quality of his situational assessment.

To date, Situation Awareness Global Assessment Technique (SAGAT) [4] is a method that is most commonly used to assess SA. The pilot SA can be assessed objectively by querying their perception, comprehension, and projection of future status, and then checking the responses with the actual situation. This allows an assessment of the stages of SA achieved by the pilot. The method requires the mission to be frozen and displays blanked out at randomly selected times. The collection of SA data at several points makes it possible to capture the changes in SA from moment to moment. However, although the method has received many validation studies [5, 6], there are some concerns in applying this method in air combat scenarios. As the method requires freezing of missions, it cannot be applied in real life situations. Even in simulated missions, the freezing can be intrusive to pilot performance and may change the nature of the task, thus violating the first requirement.

Another commonly used method is Situation Awareness Rating Technique (SART) [7]. SART is a subjective measure of SA; it requires the participants to rate the level of their SA using a Likert-type scales. The method includes three dimensions which are believed to be related to SA, i.e., attentional demand, attentional supply, and

understanding of the situation. SART is advantageous as it is relatively lower in cost and easier to administer and implement. However, the rating may be potentially confounded by workload and perceived performance rather than based on SA alone [8]. The participants may not exactly know their own awareness level, i.e., it is possible that they are not aware that they have low SA. SART rating was found to be more related to operator's confidence level of his/her SA [9]. The method itself is basically non-intrusive. However, if the ratings are to be collected several times to capture the changes in SA within one mission, a certain level of interruption to the mission cannot be avoided. Finally, it is not possible to determine the levels of SA achieved using this method.

Tham [10] developed a method to assess pilot SA specifically in air combat environment. The method is based on process tracing technique by Subject Matter Experts. It seems to be non-intrusive and able to capture the dynamic changes of pilot SA as well as the stages of SA acquired, thus fulfilling the requirements described previously. This paper aims to review this method in a greater detail to reveal its potential application in air combat missions. Several recommendations are proposed to further improve the method.

2 Process Tracing Technique to Assess Pilot SA

The tool developed by Tham is based on process tracing technique [11], which is to be completed by Subject Matter Experts (SMEs) using an observer assessment form. The participants are asked to think aloud during the mission. The process tracing can be performed concurrently during task performance, or retrospectively using a playback of the recorded scenarios. The technique tracks and monitors the SA and SA processes of the pilot, rather than measuring it quantitatively. The completed observer assessment form would indicate the progress of the different SA processes and SA achievement associated with each and every enemy aircraft, as well as that of ground threats and friendly force.

The method was applied in an experiment that aimed to examine the differences in pilot SA in performing air combat missions with and without the use of Joint Tactical Information Distribution System (JTIDS). The air combat environment was simulated using a PC-based game called Jane's USAF. The air combat scenarios were made up of multiple enemy fighter aircraft, bombers and helicopters at different ranges and headings, Surface to Air Missiles (SAMs) and Anti-Aircraft Artilleries (AAAs), as well as friendly aircrafts. Eight experienced fighter pilots (with minimum of 900 total flying hours, of which no less than 600 hours were on fighter aircraft) participated in the experiment. The assessment of pilot SA was performed by a SME, who was a fighter pilot with 1200 flying hours, 1000 hours of which were on fighter aircraft. The recorded missions were reviewed during debriefing to get the participants to further elaborate on the decisions made and maneuvers executed, as well as to fill in information gaps due to non-verbalization by the participants. The SME completed the observer assessment form through numerous reviews of the recorded simulations and explanations obtained during debriefing. Figure 1 shows an example of a completed observation sheet.

		Enemy Aircraft					
	Ranges (nm)*	**# 1**	**#2**	**#3**	**#4**	**#5**	**#6**
Knowledge of Enemy Aircraft	40-35						
	35-30						
	30-25	A: 01					
	25-20	A: 234	B: 01234	B:0123	B: 01	B: 01	
	20-15	Shot (tail)		C: 0123	C: 012	C: 012	
	15-10		Shot	C:2		C: 3	
	10-5						C: 012 Shot
	5-0						
Combat Performance	No Action						
	Kill Enemy	A: Yes	B: Yes				C: Yes
	Evade Enemy			B: Yes C: Yes	B: Yes C: Yes	B: Yes C: Yes	
	Killed by Enemy						
	Damaged by Enemy		B: Yes	B: Yes		C: Yes	
	Spike Awareness		B: Yes	B: Yes		C: Yes	
Awareness of Enemy SAM/AAA		A: Unaware of AAA, flew over it while after shooting down Mig-29 and during locking up and verifying friendly B: Unaware of SAM, flew over it while shooting down Mig-29 C: Turned back flew over SAM again					
Awareness of Friendly Aircraft		A: Lock up friendly helicopter and verified status					
Self-Awareness	Deviation (track/altitude)	No unintentional deviation					
	Positional Awareness	No problem					
	A/C Control	No problem					
Situation Description	Sit A: Single Mig-29, radar lock beaming no real threat, shot tail on, flew over AAA, Mig-29 shot down, end Sit B: First Mig-29 hot, lock up and looked at Mig-23 then came back to Mig-29 to shoot, after shot, saw Mig-21 and also Su-35 and another Mig-23, shot by mig-29 reacted to RWR and broke left cold to defend, gave up radar picture, absorbed one shot though chaff/flare end Sit C: Turned back towards east, saw Mig-23, then Mi-24, Mig-21 and Su-35, realised Mi-24 nearest at only 6.7nm shot Mi-24 and saw another 5 aircraft: Mig-21, Mig-29, Su-35, Su-27 and Mig-23all at 25-30nm, then shot at Su-35, spiked by Su-35, broke left towards north to defend seeing only spikes on RWR end							

Fig. 1. Example of a completed observation sheet

The pilot knowledge of the enemy aircraft is noted with ranges, because the distance separation between aircrafts is one of the most important factors that influence all tactical decisions, game plans and maneuvers in air combat. The level of threat posed by an enemy aircraft is usually determined by its aircraft type and weapons, particularly its radar capability and weapon effective ranges. Knowledge of critical ranges is extremely important to successfully avoid detection and weapon employment by the enemy. Ranges, together with aircraft aspect (i.e., relative heading), determine the full picture of the threat level. For instance, an enemy at 20

nautical miles (nm) coming straight would be more threatening than one beaming (going across) at 15 nm. The monitoring of ranges also parallels the scrutiny of the temporal dimension. The elapse of ranges are reflective of the time elapse, except that in air combat the effects of elapsed time can be accentuated or attenuated by the effects of relative heading.

The SME, as far as possible, defines different critical situations based on the breaks in the chain of events or occasions. For every distinct events taking place during the mission, the observer has to provide an account and relevant comments on the development of the situation. This would make the assessment more meaningful, as the development of the processes of situations and the subsequent attainment of SA can be better understood in the context of the situation(s).

The SME analyzes the behavior and verbalized thoughts of the pilots to infer quality of SA achieved. The following annotation scheme is used to note the stage of pilot's knowledge of the enemy aircraft:

- Stage 0: when an enemy aircraft first appears on radar
- Stage 1: accurate identification of existence of the enemy aircraft
- Stage 2: accurate perception of the enemy aircraft, e.g., type, range, bearing, and altitude
- Stage 3: accurate interpretation of the enemy information, e.g., threat level, if there is an opportunity of getting into offensive position
- Stage 4: accurate prediction of the next status: e.g., enemy's reaction, projection of relative position.

For combat performance, the events associated with each bandit comprise "*no action*", "*evade enemy*", "*kill enemy*", "*killed by enemy*", "*damaged by enemy*" and "*spike awareness*" (the awareness of being engaged by the enemy aircraft). These events represent all the possible options available to the pilot when dealing with bandits. The accomplishment of these combat performance events would be recorded and was annotated by "yes" and "no" for each and every situation encountered during the mission. Therefore, the assessment of the participant's combat performance is specific to the respective situation but not to that of the overall mission. Self awareness is assessed in terms of occurrence of navigational deviation and loss of awareness of own position (lost or off-track through unintentional means, whereas deliberate deviation for engagement with enemy does not constitute loss of own awareness), occurrence of loss of control of own aircraft (i.e., flying the aircraft in out-of-acceptable speed/heading/altitude that may be construed as dangerous), and occurrence of loss of awareness of wingman's position.

For instance, in Figure 1, *situation B* (the gray-shaded part), can be read as follows. For enemy aircraft #2, the knowledge of the pilot in the range of 25-20 nm was noted by "B: 01234". This indicates that situation B began with the enemy aircraft #2 appearing on the pilot's radar display in that range and that *Stage 1* through *Stage 4* knowledge level was achieved. The vertical down-pointing arrow indicates the progress of the situation with the #2 bandit closing in until it got shot in the range bracket of 15-10nm. In the process as marked by the diagonal arrow pointing to the right, bandits #3, #4 and #5 also appeared on the radar and *Stage 3* achieved on #3 while only *Stage 1* achieved on #4 and #5. From the *Combat Performance* rows, the

form indicates that #2 was destroyed but the participant also suffered damages though he was aware of the incoming missile from his positive *spike awareness*. The participant was also unaware of the SAM in situation B as indicated in the form.

3 Evaluation of the Process Tracing Technique

3.1 Advantages

An evaluation of the process tracing technique, with its associated observer assessment form in the representation format as shown, reveals its advantages as follows:

- The technique imposes minimal interruption to the task performance of the pilots. Although they were asked to think aloud during the mission, they were told to do so not at the expense of their performance. Indeed, pilots have been doing so since training days, and think aloud has become a habit. In instances when the participants did not verbalize their thoughts and actions, such information can be obtained during the debriefing by playing back the recorded mission.
- The completed form can provide an understanding of the dynamic changes of pilot's awareness and task prioritization. For instance, from the completed observation forms of the experiment [10], the following information regarding the situation assessment of fighter pilots can be obtained. When the participants encountered multiple enemy aircrafts at ranges *20nm* and beyond, *Stage 3* knowledge for all bandits can be achieved while up to *Stage 4* was obtained for the nearest bandit. The moment one to two bandits closed in within the participant's own weapon's range (around 20nm), more attention was allocated to these bandits. *Stage 3* was still maintained for the nearest and next nearest bandit but the knowledge level of the rest of the bandits had dropped to *Stage 2* or even *Stage 1*. If there were more than one bandit closing in to less than 10nm, information level of the other bandits was dropped to *Stage 2* or *Stage 1*. When one or more bandits were acquired and found to be inside *5nm*, the participant could only concentrate on one at a time to obtain a lock-on for shots. These were so-called "pop-out" bandits that had not been detected previously and had suddenly appeared and acquired by the pilot on the radarscope or JTIDS in close proximity. At such close ranges, the participant did not process information beyond *Stage 1* but only attempted to achieve lock-on and get the bandit within missile envelope (i.e., dynamic launch zone of missile or DLZ) and shoot as fast as possible. This was carried out without information processing to check for actual bearing, range and aspect angle, as they were not deemed as critical at that moment. Only after the nearest or highest-threat bandit was dealt with, he would then acquire information on others through radar search or JTIDS. The compression of time and space determined that once the enemy was within the killing zone, there was no need to "*spare extra brain bytes*" to process such information but to just get the bandit within weapon envelope and shoot.
- The technique was found to be able to reveal the differences in pilot SA when flying with and without the use of JTIDS. The differences identified were mainly

associated with *quality* (the level of completeness of the situation depiction). For instance, in a situation where there were multiple (more than two) bandits closing in from beyond 20nm, the pilots flying with JTIDS were noted to be able to achieve at least *Stage 3* and up to *Stage 4* knowledge, while they were only able to achieve up to *Stage 3* when flying without JTIDS.

3.2 Limitations

Despite of these advantages, several drawbacks of the tool were identified. First, although there was a noticeable difference in the speed of information processing and rate of SA development from initial detection (*Stage 1*) to interpretation (S*tage 3*) of bandits between operations with and without JTIDS, it was not clearly shown in the observation form. This suggests that the use of range (nm) as a reference to note down the SA progression is not sensitive enough. Second, to be valid, the judgment of SA level achieved by the pilots can only be done by SMEs who are highly experienced and have high exposure to the missions. However, this is unavoidable for a tool that is targeted to provide a specific and detailed assessment of SA. The dependency on the SME can perhaps be reduced with the use of a list of observable events and behaviors pertinent to good SA processes, such that trained observers can also perform the assessment. The achievement or non-achievement of these events can be used to indicate the level of SA attained, as well as the stages of situation assessment reached. Finally, it is difficult to quantify the overall pilot SA using this tool, which is sometimes required as an input for statistical analysis. As a result, comparing the SA level across pilots, missions, or system designs using this tool is not as straightforward as using other methods such as SAGAT and SART. Nevertheless, the method offers a deeper understanding of the processes involved in SA acquisition and maintenance, revealing the ‘hows’ rather than the ‘whats’ of SA.

3.3 Proposed Modification

Several modifications were proposed to mitigate the identified problems and to enhance the tool. An observation sheet with the proposed modifications is illustrated in Figure 2. The major modification is an inclusion of time dimension to the tool to improve the sensitivity in recording the progress in SA development. This time element also allows for easier interpretation of the data, informing what situation the pilot is facing from moment to moment, how many enemies he is exposed to at the same time frame, indicating his workload level and providing explanations to his achievement of SA. In addition, using this timeline format allows for an easier comparison of the SA progression across pilots or missions. Figure 3 used three lines to illustrate the different progressions in SA and action of the pilots. Line 1 shows that contact with the enemy is done within a shorter period of time compared to Line 2 and 3. A faster successful engagement would imply a more effective development of SA and weapon engagement by the pilot. Another situation where a pilot has a slow detection and interpretation of the enemy aircraft, which leads him to have a difficulty in gaining a point of advantage over the enemy and such that he has to evade for survival, can be translated in a similar shape to Line 3.

	Ranges (nm)	Time elapsed → 00:00	00:05	00:10	00:15	00:20
Enemy Aircraft	40-35	A: 0↓		B: 0↙ C: 0↓	B:1234	B: No Action
	35-30					
	30-25		A:1			
	25-20		A: 234		C:1	C: Damaged by C: Killed by
	20-15			A: Engage	C: Evade	
	15-10			A: Kill		
	10-5					
	5-0					
Spike Awareness			A: Yes		C: Yes	
SAM / AAA			SAM: Damaged by			
Friendly Aircraft						Loss awareness of wingman
Deviation (track)		On track	On track	Deviate	Deviate	Deviate
Deviation (altitude)		On track	On track	On track	Deviate	Deviate
Remarks			Unaware of SAM missile		Aware that enemy B is not a threat, no action is necessary	

Fig. 2. A simple illustration to the proposed modification

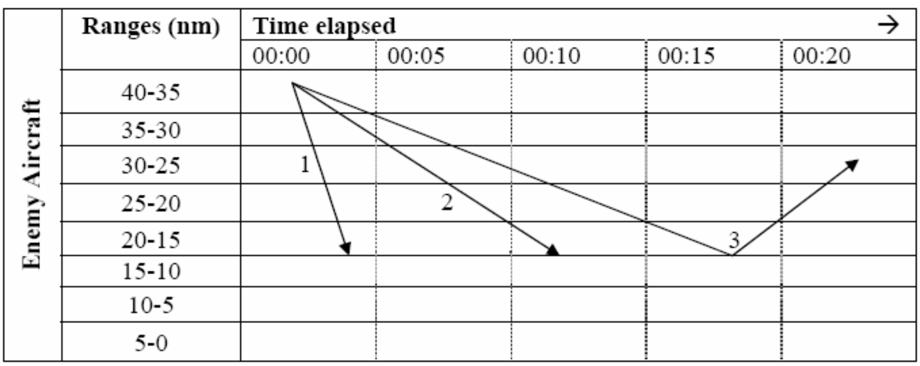


Fig. 3. Illustration of different progression in SA and action

Some modifications to the notation are also proposed, i.e., using alphabets to represent the enemy aircrafts instead of situations, and incorporating the actions taken to deal with the enemy within the time frame instead of on separate rows. Another additional feature proposed is to note the direction of enemy approaching. As mentioned previously, a combination of ranges and relative heading can indicate the threat level of the enemy aircraft. For instance, Figure 2 indicates that enemies B and

C are approaching at the same ranges, but B is beaming while C is coming straight. Thus, the pilot who understands that B has lower threat would decide to engage C first and perhaps ignore B if time and resources do no permit a successful engagement with B. As all of these are shown in the observation sheet, we can understand that "no action" taken on enemy B does not mean that the pilot had a low SA on B, but that he had a good picture of the overall situation and an effective management of resources.

4 Conclusion

Air combat environment is highly dynamic. A tool to assess pilot SA that does not alter this dynamic nature is desired. In this study, a measurement tool based on process tracing technique was evaluated. The tool was found to be powerful to elicit the situation assessment strategies and processes, produce comprehensive and detailed information on pilot SA, and trace the dynamic changes in SA from moment to moment. Based on this information, further investigation can be done to find out the source of SA decrement and the behavior that can help to recover loss of SA. The tool is thus viewed to have a good potential to be applied in air combat missions. However, some problems with the method were also noted, and modifications to minimize these problems were proposed. However, further studies to evaluate and validate this modified tool are required.

References

1. Endsley, M.R.: Situation awareness in aviation systems. In: Garland, D.J, Wise, J.A., Hopkin, V.D. (eds.) Handbook of Aviation Human Factors, pp. 257–276. Lawrence Erlbaum Associates, Mahwah (1999)
2. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Human Factors 37, 32–64 (1995)
3. Carreta, T.R., Perry, J.D.C., Ree, M.J.: Prediction of Situational Awareness in F-15 Pilots. The. International Journal of Aviation Psychology 6, 21–41 (1996)
4. Endsley, M.R.: Situation awareness global assessment technique (SAGAT). In: Proceedings of the IEEE National Aerospace Electronic Conference, vol. 3, pp. 789–795 (1988)
5. Endsley, M.R.: Predictive utility of an objective measure of situation awareness. In: Proceedings of the Human Factors Society 34th Annual Meeting. Orlando Florida, pp. 41–45 (1990)
6. Endsley, M.R.: Direct measurement of situation awareness: Validity and use of SAGAT. In: Endsley, M.R., Garland, D.J (eds.) Situation Awareness Analysis and Measurement, pp. 147–173. Lawrence Erlbaum Associates, New Jersey (2000)
7. Taylor, R.M.: Situational awareness rating technique (SART): the development of a tool for aircrew systems design. In: Situational Awareness in Aerospace Operation (AGARD-CP-478). NATO-AGARD, Neuilly Sur Seine, France (1990)
8. Jones, D.G.: Subjective measures of situation awareness. In: Endsley, M.R., Garland, D.J (eds.) Situation Awareness Analysis and Measurement, pp. 113–128. Lawrence Erlbaum Associates, New Jersey (2000)

9. Endsley, M.R., Selcon, S.J., Hardiman, T.D., Croft, D.G.: A comparative analysis of SAGAT and SART for evaluations of situation awareness. In: Proceedings of the 42nd Annual Meeting of the Human Factors and Ergonomics Society. HFES, Chicago IL, pp. 82–86 (1998)
10. Tham, Y.M.: A study on the assessment of situational awareness of pilots operating with and without tactical data link systems in simulated air combat. Unpublished Final Year Report. School of Mechanical and Production Engineering, Nanyang Technological University, Singapore (2003)
11. Hollnagel, E., Pederson, O., Rasmussen, J.: Notes on Human Performance Analysis. Riso National Laboratory, Electronic Department, Roskilde Denmark (1981)

Analysis of Human Factors Integration Aspects for Aviation Accidents and Incidents

Sun Ruishan, Wang Lei, and Zhang Ling

Research Institute of Civil Aviation Safety
Civil Aviation University of China
300300, Tianjin, China
rssun@cauc.edu.cn, sunrsh@hotmail.com

Abstract. Aviation accidents have been contributed mostly by human factors since commercial flight. So it is a key of improving aviation safety to analyze accidents, incidents and other occurrence through human factors model and make preventing measure. The paper presents an analysis method named EEAM based human factors integration.

Keyword: Human Factors; Aviation Safety; SHEL; HFACS; EEAM.

1 Introduction

The human element is a causal or contributing factor in the great majority of aircraft accidents and incidents. Interestingly, the ratio of human factors accidents is relatively constant across the entire aviation spectrum. This percentage (generally accepted at 60% to 80%) does not vary with pilot experience or types of aircraft flown.

Human factors methods and tools are seldom used in the analysis of accidents/incidents database. What we required is a general analysis framework which can find the deep and hidden reason behind accidents.

This paper tried to put forward an improved method to analyze the human factors integration aspects in accidents/incidents, then applied it into database analysis and got several constructive conclusions and suggestions.

2 General Model of Human Factors Analysis

There have been many research theories and products in the field of human factors research, such as SHEL Model, Reason’s Swiss Cheese Model and HFACS. Some of them have been applied in aviation very well and improved the aviation safety greatly.

SHEL Model is a conceptual model of human factors and named after the initial letters of its components: Software (procedures, etc.); Hardware (machine); Environment (operational and ambient); Liveware (human element). SHEL Model was developed first in 1972 (Edwards, 1972), and then was improved as “building block” model used today by Hawkins (1984).

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 834–841, 2007.

In the SHEL model, the match or mismatch of the blocks (interface) is just as important as the characteristics described by the blocks themselves. These blocks may be re-arranged as required to describe the system. A connection between blocks indicates an interface between the two elements. Each element of the system should be described both functionally and physically if possible.

Another famous approach to the genesis of human error is proposed by James Reason (1990). Generally referred to as the “Swiss cheese” model of human error, Reason describes four levels of human failure, each influencing the next. He firstly put forward the concepts of active failures and latent failures.

Working backwards in time from the accident, the first level depicts those Unsafe Acts of Operators that ultimately led to the accident. More commonly referred to in aviation as aircrew/pilot error, this level is where most accident investigations have focused their efforts and consequently, where most causal factors are uncovered.

HFACS (Human Factors Analysis and Classification System) was developed as a taxonomic system to categorize both the latent and active causal factors that have been identified in aviation accidents. Its purpose is to provide a framework for use in aviation accident investigations and as a tool for assessing accident trends.

Drawing upon Reason’s (1990) concept of latent and active failures, HFACS describes four levels of failure:1) Unsafe Acts, 2) Preconditions for Unsafe Acts, 3) Unsafe Supervision, and 4) Organizational Influences. The HFACS framework bridges the gap between theory and practice by providing investigators with a comprehensive, user-friendly tool for identifying and classifying the human causes of aviation accidents.

Through analyzing the SHEL and Swiss cheese Model, we could find that the accident was always caused by multi-aspects of reasons in different levels. If we cut down one link of the events chain, the accident will be prevented. The HFACS method is an application in practice; it can help us to find out the “holes in the cheese”, and then establish the prevention measures of accidents. The challenge for accident investigators and analysts is how best to identify the causal sequence of events, in particular that 60-80 % associated with human error. Referring to the human factors research in China, we should study the basic theories steadfastly, and then develop a corresponding application model suiting our aviation industry to improve the safety named EEAM.

3 Improved Elementary Events Analysis Method (EEAM)

As early as 1996, Civil Aviation Safety Research Institute of CAUC has developed a method called as “Elementary Events Analysis Method” (EEAM). By EEAM all occurrences including accidents, incidents, human errors, mechanical faults and weather problems can be analyzed under the same frame. All data collected by different program are united and accumulated. So a uniform database for airman, crew, fleet, airline, region, and whole country can be built up. Based on previous products, we will propose the improved EEAM for being applied more easily and effectively.

3.1 Improved EEAM

Firstly, we should give the definition of elementary events. Elementary Events are understood as such an occurrence which is a single human error or a mechanical fault or an environment problem, and it is a link of the events chain, and an accident can be prevented by removing it. So maybe one accident contains several elementary events, the analyzing model was given like figure 1.

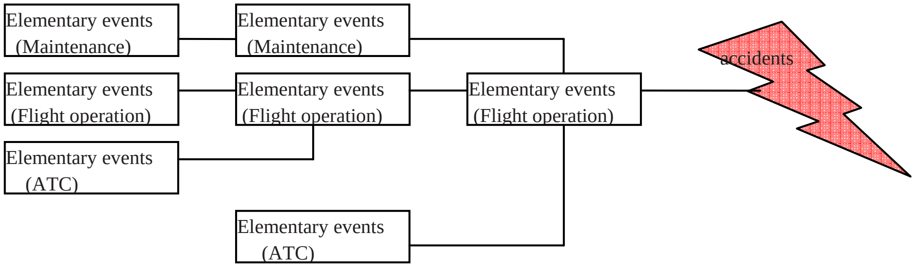


Fig. 1. EEAM analyzing model

Then, the improved EEAM provides an analysis frame to standardize the events analysis and build up a united database which will be greatly helpful for looking into the interior of civil aviation and finding the deficiencies of a system, the frame is as the following:

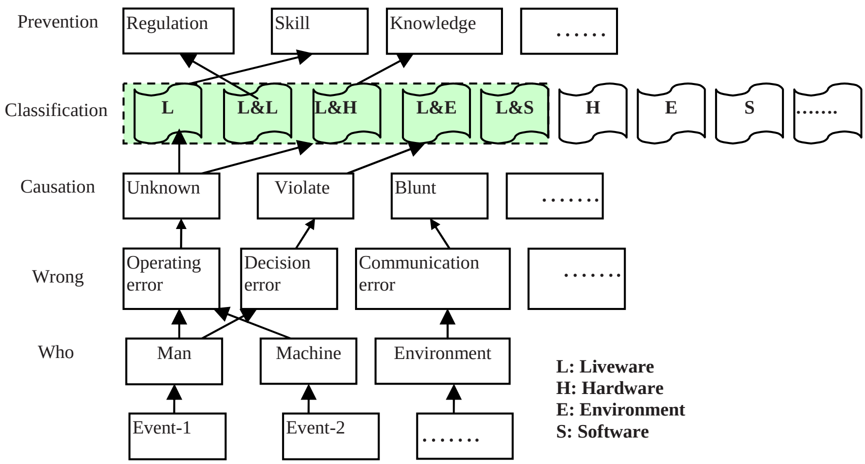


Fig. 2. Improved EEAM Analysis Frame

By this improved EEAM all occurrences including accidents, incidents, human errors, mechanical faults and weather problems can be analyzed under the same frame. All data collected by different program are united and accumulated.

Compared with the previous EEAM, the level of Classification was added in the improved EEAM for analyzing the human factors integration aspects more reasonably and directly. In this level, we refers to the classification method of SHEL Model, classifies the causations as 4 main types of live ware, hardware, environment and software, we focus on analyzing the live ware and its relations with other elements in this paper.

The database we developed includes three parts: the pilot's database, the Air Traffic Controller's database and the maintenance personnel database. We have analyzed the accidents, incidents and human errors by using the improved EEAM.

3.2 Analysis Process

Now we analyze one accident by using this improved frame. At second July, 1999, a helicopter (B-7953) got in GPS malfunction after taking off, but the crew neither detected it, nor utilized other navigation tools. Finally the helicopter dropped into the sea due to lack of oil for flying to standby airport.

According to the improved EEAM model: Firstly, we defined the reasons of this accident are crew factors; Secondly, we classified the error types in this accident as : (1) inefficient observation; (2) neglecting ; (3) communication problem; Thirdly, we concluded the causing as: ① depending on automatic blindly (crew did not monitor the flight effectively) ; ② crew had not prepared sufficiently (they neither got correct weather information, nor evaluate the flying plan); ③ crew did not cooperate harmoniously and communicated effectively; Fourthly, we classified the causing as following: the ① was human machine integration error, ② was human-self error and ③ was human integration error.

Finally, we should make some corresponding prevention measures according to the three types of factors. For this accident, firstly crew should have a scientific standpoint of automatic using, and know how and when to use it. Secondly, we should improve the crew resource management (CRM), improve their safety consciousness, and require them to get sufficient preparation before every flight. Thirdly, it is also very important to create a harmonious environment for crew, let them communicate and cooperate in a good way.

3.3 Characters

It is obvious that the improved Elementary Events Analysis Method is a useful tool for accident analysis and prevention. The technique is characterized by the following: the level of Classification was added for analyzing the human factors integration aspects more reasonably and directly, it combines with the advantages of HFACS and SHEL Model. Accident, incident, and other safety occurrence, all are analyzed under the same frame.

Human factors are put into the centre of safety management. This is a main progress made by aviation community in recent years. It focuses on unsafe actions and their prevention measures which are treated in a systematic way.

4 Analysis of Accidents/Incidents by Using EEAM

4.1 Analysis of Incidents During 1989-2005

We analyzed 2021 cases of incidents of Chinese civil aviation industry during 1989-2005 and got the some valuable information by EEAM.

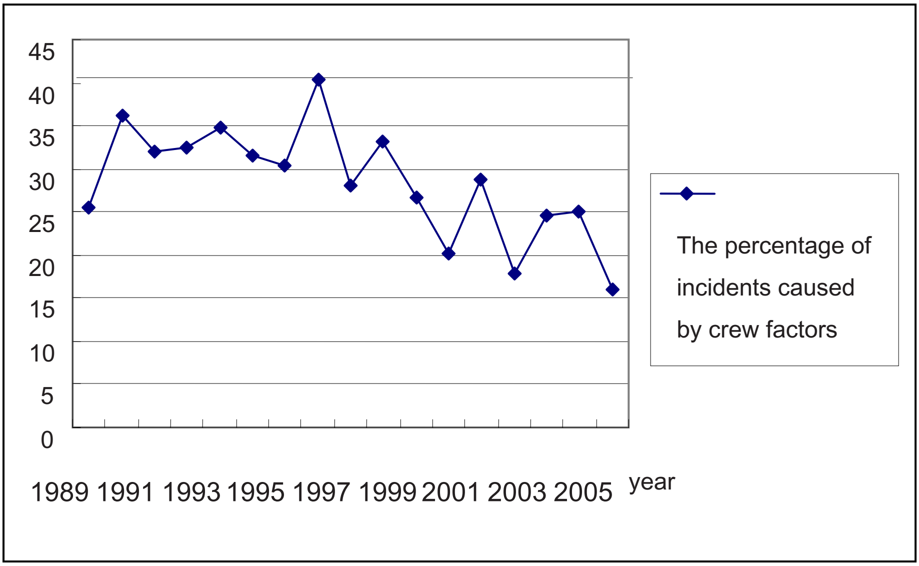


Fig. 3. The percentage of incidents contributed by crew (1989-2005)

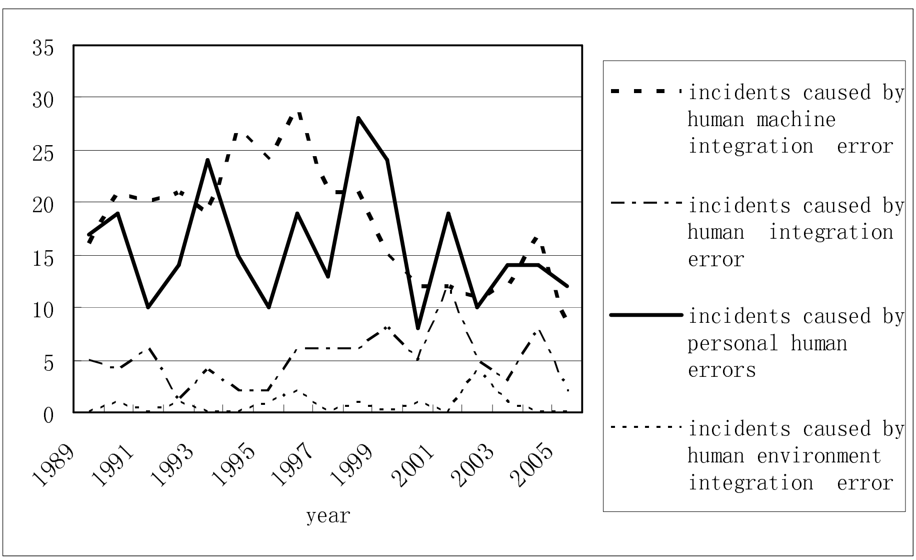


Fig. 4. Incidents involved in human factors (1989-2005)

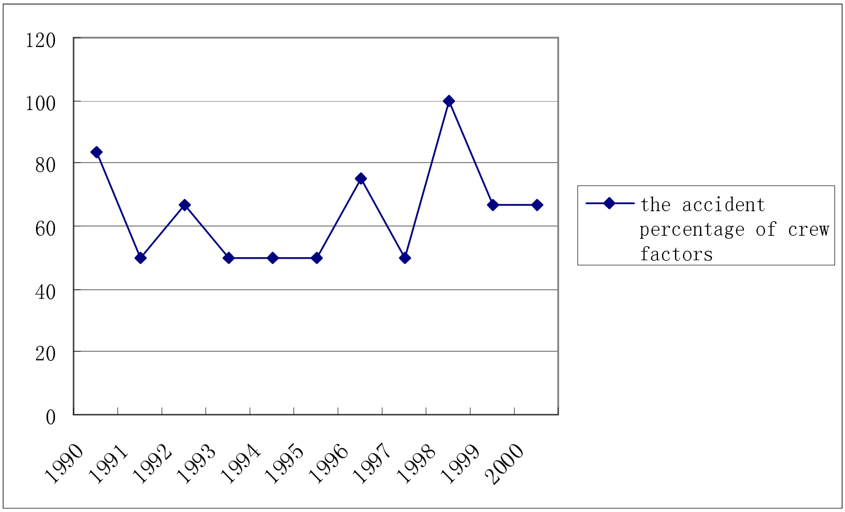


Fig. 5. The percentage of accidents caused by crew factors (1990-2000)

As seeing from figure 4 and 5, the trend of incidents caused by crew factors is becoming declining but very slowly. Meanwhile, human machine integration errors and personal human errors that account for the majority of human errors affect the happening of incidents greatly. In addition, there is an alternating tendency existing between them.

4.2 Analysis of Accidents During 1990-2000

We also made a statistic and analysis of 53 cases of accidents by using this improved model. The result shows that there are 33 pieces of accidents induced by crew factors.

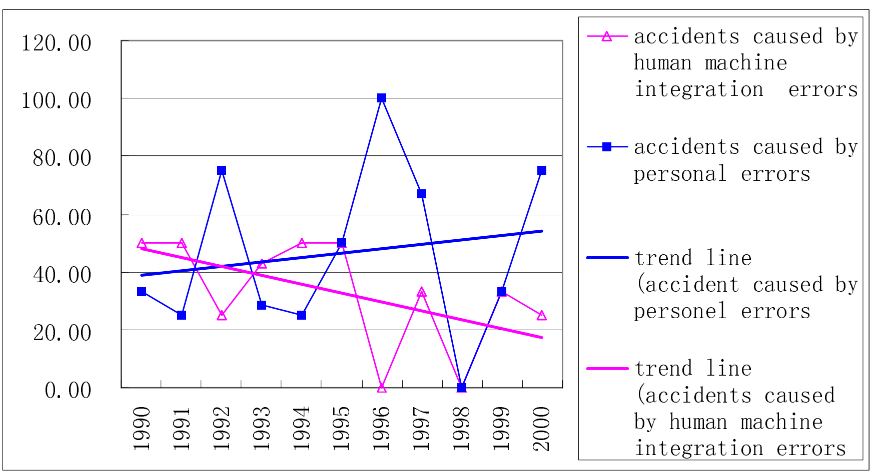


Fig. 6. Accidents resulting from personal errors and human machine integration errors (1990-2000)

4.3 Contrastive Analysis Between Incidents and Accidents

The analysis results above show that human factors make a different influence in accidents and incidents. During the 17 years of 1989-2005, the incident caused by crew factors is 28%; but this number is 63% for the accidents. Furthermore, the influence degree varies in the four types of human errors, as the result following:

Table 1. The influence of Crew factors and other four types of human errors for incidents and accidents

	The influence degree of these to crew factors (%)				The influence of crew factors (%)
	Human machine integration errors	Human integration errors	Personal human errors	Human environment integration errors	
Incidents	53.4	14.8	47.1	2.1	28
Accidents	48.5	21.2	60.6	3	62

We concluded that human machine integration errors have caused most of incidents, but personal human errors for most of accidents. However, they are both very important as the main constitutes of human factors, so we should pay more attention on them.

5 Conclusions

The EEAM based on fault tree analysis (FTA), SHEL model and Swiss Cheese model has an apparent advantage. The EEAM may analyze accident, incident, occurrence, and human factors event by unitive frame. Especially, a safety manager and other administrator may get consultation of improving aviation safety. Primary use for analysis of accidents and incidents has revealed the EEAM a useful and powerful tool. It will get some valuable result of safety management to analyze all occurrences deeply and detailed

References

1. Hanhui, L., Ruishan, S.: Events Analysis Technique. In: Safety Science Conference (1999a)
2. Hanhui, L., Ruishan, S., Xiushan, Z.: Elementary Events Analysis Method. Journal of Civil Aviation University of China 15(3) (1997)
3. Jones, A.P.: Climate and measurement of consensus: A discussion of organizational climate. In: Cole, S.G., Demaree, R.G., Curtis, W. (eds.) Applications of Interactionist Psychology: Essays in Honor of Saul B. Sells, pp. 283–290. Earlbaum, Hillsdale (1988)
4. Murray, S.R.: Deliberate decision making by aircraft pilots: A simple reminder to avoid decision making under panic. The International Journal of Aviation Psychology 7, 83–100 (1997)

5. Ruishan, S.: Aviation Accidents Analysis and Prevention Strategy, vol. 3. Safety Science and Technology Science Press, Beijing/New York (2002)
6. Muchinsky, P.M.: Psychology applied to work, 5th edn. Brooks/Cole Publishing Co, Pacific Grove, CA (1997)
7. National Transportation Safety Board, Air Florida, Inc., Boeing 737-222, N62AF, Collision with 14th Street bridge, near Washington National Airport, Washington, D.C. (January 13, 1982) (Tech. Report NTSB-AAR-82-8). Washington: National Transportation Safety Board (1982)
8. Orasanu, J.M.: Decision-making in the cockpit. In: Wiener, E.L., Kanki, B.G., Helmreich, R.L. (eds.) Cockpit resource management, pp. 137–172. Academic Press, San Diego (1993)
9. Reason, J.: Human error. Cambridge University Press, New York (1990)
10. Shappell, S.A., Wiegmann, D.A.: Human error in commercial and corporate aviation: An analysis of FAR Part 121 and 135 mishaps using HFACS. Aviation, Space, and Environmental Medicine, 70, 407 (1999a)

Development and Evaluation of a Multimodal Touchpad for Advanced In-Vehicle Systems

Roman Vilimek[1] and Alf Zimmer[2]

[1] Siemens AG, Corporate Technology, User Interface Design
Otto-Hahn-Ring 6, 81730 Munich, Germany
roman.vilimek.ext@siemens.com
[2] University of Regensburg, Department of Experimental Psychology
93053 Regensburg, Germany
alf.zimmer@psychologie.uni-regensburg.de

Abstract. Multimodal interaction can substantially improve human-computer interaction by employing multiple perceptual channels. We report on the development and evaluation of a touchpad with auditory, tactile and visual feedback for in-vehicle applications. In a simulator study, we assessed its suitability for interacting with a menu-based on-board system and investigated the effects of uni-, bi- and trimodal feedback on task and driving performance, workload and visual distraction in comparison to a conventional rotary push-button. In summary our results show that users clearly benefit from additional non-visual feedback while driving. When using the touchpad with multimodal feedback, our subjects also reached a higher level of performance compared to the rotary push-button.

Keywords: Multimodal, haptics, tactile feedback, auditory feedback, driving.

1 Motivation

In recent years, the automotive industry had to face the challenge of integrating a huge number of additional functions for driver information, communication and navigation in modern cars. To avoid an excessive amount of controls and displays, a growing number of car manufacturers use a computer-like menu-based system with a display in the center console and a central control for manual input. This central control is in most cases a form of rotary push-button positioned between the front seats. As an alternative, we propose a touchpad-based central control with multimodal feedback to overcome some inherent drawbacks of the rotary push-button concept.

Using a touchpad to control the cursor movements offers several advantages. Menu items or interaction elements for setting functional parameters are typically not displayed in a circular layout. When using a rotary push-button, the user is required to convert the circular input movements to the intended linear movements of the cursor. This can easily get confusing when vertical arrangements of interaction elements like menus and horizontal arrangements typically used for sliders are intermixed. Furthermore, the rotary push-button does not allow for two-dimensional movements on the screen. Taken together, the rotary push-button limits the systematic application of

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 842–851, 2007.

natural mappings [1], one of the basic requirements for intuitive man-machine interaction. With a touchpad, cursor movement is simply controlled linearly in two dimensions by moving the finger in the desired horizontal or vertical direction.

However, a conventional touchpad is most likely not suitable for in-vehicle use as actions are only visible on the graphical user interface. To minimize visual distraction we built a touchpad with additional auditory and / or haptic feedback. Any event shown visually is also displayed redundantly in a non-visual modality. Auditory feedback is given by playing short auditory icons, i.e. natural sounds which are mapped to system events by analogy (e.g., click-sounds; for details see [2]). Appropriate freely programmable actuators are used to generate vibrotactile feedback in form of vibration patterns underneath the touchpad. If, for instance, the user moves the selection cursor from one menu element to the next, he hears a clicking sound and feels a short tactile impulse at his fingertip. Thus, the visually displayed selection of a new element becomes palpable and audible. When a menu item is activated, another sound or vibration pattern is played. Characteristics of the prototype are described below in more detail.

2 Research Goals

The goals of our study were twofold: 1) On a theoretical level we explored the effects of multimodal feedback, especially the relative contributions of auditory and haptic feedback in direct comparison; 2) On a practical level we strove for a more ergonomic solution to the problem of manual input for in-vehicle secondary tasks while driving, by providing the users with a touchpad instead of a rotary push-button.

According to Wickens' multiple resource model [3], an additional auditory display improves time-sharing performance in a multiple-task setting with high visual load. This prediction has been confirmed in various contexts (e.g. [4, 5]). Using a touchpad with additional auditory, tactile or auditory plus tactile output should thus allow the driver to perceive more driving-relevant information while interacting with the on-board system compared to a touchpad with visual feedback only. Furthermore it has been shown that ballistic targeting tasks can be accomplished faster with additional feedback [6, 7]. However, several aspects need closer analysis.

Whereas there is substantial evidence of the advantageous effects of multimodal feedback in general, there is no universal multimodal theory that allows the deduction of success probabilities for a novel interface in a concrete task environment. The multiple resource model provides a general framework for theoretical considerations, but in practice too many factors influence the interaction between users and a particular novel interface solution. Thus, the anticipated positive effects of multiple feedback modalities for a touchpad-based central control for on-board systems need experimental confirmation.

Based upon previous research it seems very likely that auditory feedback is an appropriate and accepted way of providing additional information to cursor movements. However, it is unclear to what extent users will benefit from vibrotactile information. Tactile displays are not new, but have mainly been used for highly specialized applications such as displays for blind people. However, as the cutaneous sense is still under-utilized in human-computer interaction [8] there is not sufficient knowledge on

how to design vibrotactile patterns as a means of informing the user on the success of intended interaction events. Systematic research on the use of structured, abstract vibrotactile patterns to communicate messages non-visually (so-called *Tactons* or tactile icons) has just started in the recent past [9, 10]. The present study examines whether Tactons are as effective as auditory messages in providing non-visual action feedback.

As an external baseline we included a rotary push-button in our study. We used a standard conventional rotary push-button switch, i.e. menu items are selected by turning the switch, and the selection is confirmed or activated by pressing it. There was no advanced auditory or haptic feedback for this input device, but in addition to the visual feedback on the graphical user interface (GUI) the mechanical properties of the switch enabled the users to infer important information non-visually. For instance, when turning the switch they feel the detent and hear a clicking sound when it snaps into position. Our subjects had to accomplish several experimental tasks using either this rotary push-button or the touchpad with unimodal (visual only), bimodal (visual+tactile or visual+auditory) and trimodal (visual+tactile+auditory) feedback. This experimental setup allows for direct comparisons between the different input devices and the effects of multimodal feedback.

3 The Multimodal Touchpad

The multimodal touchpad was evaluated in a cockpit mock-up. It was placed between the two seats and replaced the rotary push-button as central input device. In our experimental tasks, subjects had to accomplish several tasks which are typical for on-board systems. If, for instance, they wanted to use entertainment functions, they first had to select the menu entry “entertainment” in the main menu by moving their finger up or down in order to place the cursor onto the corresponding element. A short tap on the touchpad then activated the selection. Moving the cursor and clicking on the touchpad are similar to the actions performed with an ordinary laptop touchpad. However, the cursor did not move freely like a mouse cursor. It was used as a selection cursor, visible on the GUI as a highlighted frame around a menu item or any other selectable control element and restricted to move on these elements. In addition to the visual indication of movement, tactile feedback for the touchpad was realized by using unbalanced electric motor vibrators and a tactile transducer (bass shaker) mounted on a plate below the touchpad and a wrist rest. The wrist rest supported the users arm while he or she performed input operations. Different forms of tactile impulses were designed for output at the user’s index finger and wrist. Figure 1 shows the components of the touchpad prototype and its integration into the cockpit mock-up which was used for the driving simulation. It also shows the TFT display next to the steering wheel for visual output and the loudspeaker behind the displays for auditory feedback.

The tactile transducer was a Fischer Amps shaker (type “Bass Pump III”) specialized for low frequency vibrations, which was amplified with an additional low frequency amplifier (AURA Interactor) and connected to one of the soundcards of a PC. It was positioned right under the touchpad. The unbalanced electric motor vibrators under the wrist rest were those of a Logitech® Rumblepad™ 2, which was connected

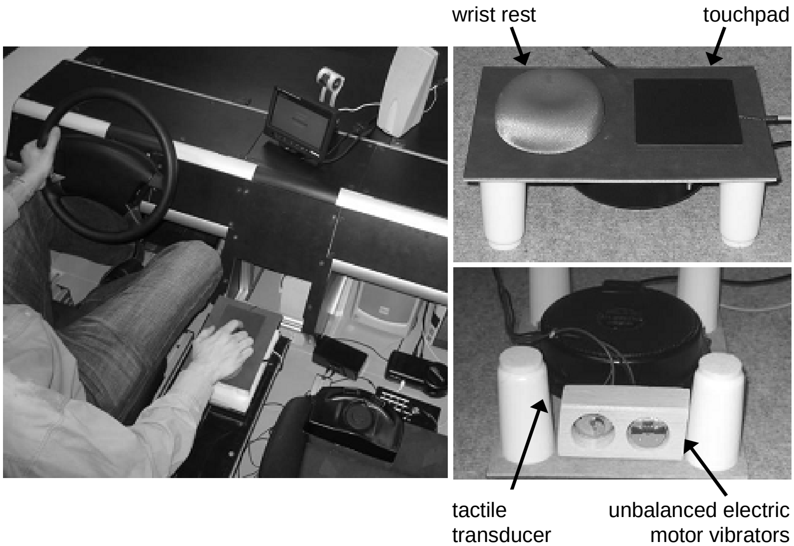


Fig. 1. Left side: Cockpit mock-up. Right side: Constituent parts of the multimodal touchpad.

to the PC via USB. The motors were generally used for slower but smoother feedback at the wrist whereas the tactile transducer generated single, short impulses at the fingertip. The tactile feedback events were designed either with the software Immersion Studio® by Immersion, Corp. (for the motors) or consisted of simple low frequency sine waves generated with the help of the sampling software Cool Edit 2000 (Syntrillium Software Corporation).

Any feedback event shown visually was also implemented in the non-visual modalities. In our experiment, several different types of events had to be designed. Most important are the selection and the confirmation/activation of an item, which are described here in detail. A movement of the selection cursor was signaled by a 20 ms 50 Hz sinus impulse issued by the tactile transducer. The auditory feedback of this event was a metallic clicking sound (15 ms). The confirmation/activation of a selected element by tapping on the touchpad was represented auditorily with a deep knocking sound, like knocking on a table (60 ms). Tactile feedback consisted of a 20 ms 30 Hz sinus wave and additional activity of the motor vibrators for 150 ms. As the index finger had to be lifted off the touchpad during tapping, the tactile feedback was primarily perceived as a vibration of the wrist rest.

Considerable care was taken to ensure that tactile events were well perceivable but nevertheless did not produce audible artifacts. The frequency range was restricted to very low frequencies. The plate with the touchpad and the tactile actuators was placed on a special pad which absorbed the structure-borne sound. Thus, the effects of tactile feedback were only marginally contaminated with effects of auditory perception.

4 Methods

4.1 Participants

Twenty participants (twelve female, eight male) took part in the present study. They were monetarily compensated or received a small gift for their participation. Their ages ranged from 22 to 36 years. All had held a valid driving license for at least two years and drove a car regularly. This was essential because the simulated driving task had to be dealt with and prioritized in the same way as real driving.

4.2 Apparatus and Driving Simulation

The study took place in the Corporate Technology usability lab at Siemens AG. The subject's workplace during the experiment was the driver's seat of a fixed-base cockpit mock-up. Depending on the experimental condition, the touchpad or the rotary push-button was positioned between the seats to the right of the driver. The experiment was run on a Fujitsu Siemens PC with 3.4 GHz CPU equipped with two identical ESS Technologies ESS1969 Soundcards (one auditory, one for vibrotactile feedback). The experimental conditions were controlled by a Microsoft® Visual Basic® program that generated the interface, processed the user inputs, controlled the multimodal feedback and logged all user actions. The central display (7" TFT running at a resolution of 800 × 480) of the simulated on-board system was placed on the dashboard to the right of the steering wheel. Figure 2 shows the setup with either the touchpad or the rotary push-button. Auditory feedback was issued by a loudspeaker behind the display.

Fig. 2. Experimental setup with multimodal touchpad (left) and rotary push-button (right)

For the baseline comparison, a rotary push-button switch was vertically mounted in a case (ddm hopt+schuler, type 427; 30 positions, 360° endless, operating torque: 1.5 Ncm, actuation force: 6 N, detent angle: 12.5°). The shaft was covered with a round plastic cap (approximately 4 cm in outside diameter).

To assess visual distraction a digital video camera (Philips SPC600NC) connected to a separate Fujitsu Siemens PC with 2.8 GHz CPU recorded the subjects' face (30

frames/second; resolution: 640 × 480). After each experimental session, the video was analyzed manually in slow motion using a customized video tool.

The driving simulation used in this experiment was the Lane Change Task (LCT) [11]. The driving task consists of lane keeping and lane changing maneuvers in response to traffic signs at the side of the road. By comparing the driving performance with a normative model, the LCT allows the measurement of distraction caused by in-vehicle systems. It delivers a deviation score which covers important aspects of driving like the quality of maneuvers and the perception of driving-relevant information.

4.3 Tasks

The experimental tasks represented typical aspects of input operations with in-vehicle systems. The participants had to adjust numerical values for temperature or volume using horizontal or vertical sliders, select characters from a 2 dimensional array to enter a navigation destination, select items from a list and browse through different menus. In half of the experimental trials they had to drive concurrently. They were instructed to prioritize driving and treat it strictly as the primary task.

4.4 Design

This study was a 2 × 5 factorial, within-subject design. The independent variables were *task situation* and type of *HMI*. Task completion time, subjective mental workload (RTLX, a modified version of the NASA TLX [12]), driving performance (LCT) and display viewing times were the dependent variables.

All subjects interacted with the on-board system either in a single-task static test situation involving only the input tasks or in a dual-task situation with concurrent driving simulation (=independent variable: task situation). Furthermore, all subjects used all different HMI variants available in this study, as defined by the type of interaction device (touchpad or rotary push-button) and the type of feedback given when using the touchpad. Visual feedback was always present. Thus, the second independent variable, HMI, had five levels:

- **TPv:** Unimodal touchpad with only visual output without additional feedback
- **TPva:** Bimodal touchpad with visual and auditory feedback
- **TPvt:** Bimodal touchpad with visual and tactile feedback
- **TPvta:** Trimodal touchpad with visual, tactile and auditory feedback
- **RPB:** Rotary push-button switch

The subjects first completed all experimental tasks in five blocks, defined by the type of HMI used. A Latin square was used to counterbalance the block sequences with a new random sequence after any group of five subjects. In each block the experimental tasks were performed first without any driving task and after that with concurrent simulated driving. Additionally, all participants had to complete one LCT run without simultaneously using the on-board system to deliver a baseline measure of their driving performance. Subjects performed this task either at the beginning, in the middle or at the end of the experiment.

4.5 Procedure

Before engaging in the experiment, the participants were pre-screened to ensure that they met the qualification criteria. The experimenter explained the experimental tasks and the driving task. The participants received a short driving training session. They also completed several practice runs with the on-board system.

Each experimental block started with a short practice run on the current HMI. After finishing all tasks with and without driving, the participants filled out the workload questionnaire at the end of each block. Finally, after finishing all experimental blocks and the baseline LCT, the participants were encouraged to report their opinions and preferences regarding the HMI variants.

5 Hypotheses

On the one hand, the better the natural mapping between cursor movements and GUI responses, the greater the likelihood that the touchpad will lead to more intuitive and thus faster interaction. On the other hand, this advantage may be substantially reduced by the increased visual load if the touchpad does not provide additional non-visual feedback. However, when the subject focuses all visual and cognitive resources on the single-task situation without driving, only marginal differences are expected between the rotary push-button and the touchpad, independently of additional feedback. But as soon as a primary visual task, the LCT, is present, the multiple resource model [3] implies a stronger decrease in performance for the secondary task if only visual presentation is available. Thus, the touchpad variants with bi- or trimodal feedback should have a general advantage over the unimodal visual-only touchpad. Furthermore it is very likely that the rotary push-button will also lead to better time-sharing than the unimodal touchpad because of its inherent mechanical feel and audible feedback. Considering basic research results [13], it also seems likely that any positive effects will be more pronounced with trimodal than with bimodal feedback.

6 Results and Discussion

Data collected in the present experiment were analyzed using inferential statistics (ANOVA). The alpha level for significance was chosen to be .05.

Regarding task completion times, no major differences were found between the different HMIs (TPv, TPva, TPvt, TPvta, and RPB) in the single-task situation. The values ranged between 1.97 and 2.07 s. But when the subjects had to drive concurrently in the dual-task situation, a different picture emerged (see Fig. 3). Of course it took more time to fulfill the tasks regardless of the HMI used (significant main effect for task situation: $F(1,19) = 34.86$, $p < .001$). More interesting are the differences between the HMI variants while driving. If the subjects had only visual feedback when interacting with the touchpad they needed 3.41 s. This is considerably longer than the mean task completion times for the bi- or trimodal touchpads, TPva: 2.82 s, TPvt: 2.86 s, TPvta: 2.96 s. Completing a task with the rotary push-button took the participants almost as long as with the unimodal touchpad (3.39 s).

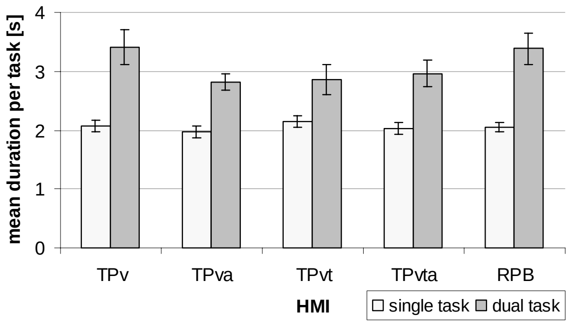


Fig. 3. Task completion times for each HMI variant in single-task (static test) and dual-task (with simulated driving) situations. Error bars indicate standard error of the mean.

A significant interaction between HMI and task situation ($F(4,76) = 2.84$, $p = .03$) confirms the hypothesis that the differences between the HMI variants show up only if there is a resource conflict in the dual-task situation. If additional feedback is available for the touchpad, task completion times are clearly shorter. Interestingly, there is no indication that trimodal feedback has a more pronounced effect than bimodal. Furthermore the relatively long task completion times of the baseline RPB are worth mentioning. Using the RPB took our subjects almost as long as using the unimodal touchpad TPv. Based upon these data, even a conventional touchpad with only visual feedback seems to be a reasonable alternative.

An analysis of the LCT data reveals that the subjects successfully prioritized the primary task: safe driving. The deviation from the normative model (optimal lane keeping and changing) was very similar in the baseline run without any additional task and when subjects were driving under dual-task conditions, i.e. the quotient of the deviation with/without secondary task is near 1. For each HMI, the average values are: TPv: 1.00, TPva: 1.00, TPvt: 1.03, TPvta: 1.05, RPB: 1.04. A one-factor repeated measures ANOVA did not indicate significant differences ($F(4,76) = 1.7$, $p = .15$).

To assess visual distraction, we analyzed the duration of gazes at the display during driving. The longest mean duration of a single gaze at the display (see Fig. 4, A) was found when subjects were interacting with the unimodal touchpad (TPv: 0.79 s), whereas the auditory touchpad produced the shortest control gazes (TPva: 0.66 s). Accordingly, the remaining touchpad variants and the RPB fall within this range (TPvt: 0.73 s, TPvta: 0.70 s, RPB: 0.67 s). A one-factor ANOVA evaluated this effect of HMI and confirmed it as significant ($F(4,76) = 2.46$, $p = .05$). However, the differences between the HMI variants are rather small, so there may be doubt as to their practical impact. Taking account of the total time the subjects spent looking at the display (see Fig. 4, B) helps to evaluate the consequences of longer input operations.

Major differences between the HMI variants are obvious at first sight. They proved to be statistically meaningful ($F(4,76) = 11.80$, $p < .001$). These data provide evidence that the unimodal touchpad (TPv: 19.15 s) leads to noticeably prolonged visual distraction compared to the baseline RPB (16.75 s), which limits its usefulness in practice. They also suggest a clear advantage of the bi- and trimodal touchpads over the baseline RPB (TPva: 11.84 s, TPvt: 14.20 s, TPvta: 13.35 s).

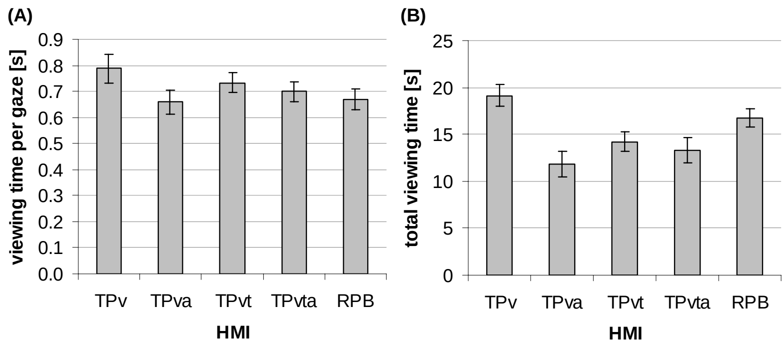


Fig. 4. A) Average duration of control gazes at the display and B) total viewing time accumulated over all experimental tasks for each HMI variant (with standard errors)

The subjective workload ratings deliver a similar pattern of results. The overall score of the RTLX ranges from 0 to 100, with higher values indicating a higher workload. Subjects reported the highest workload when using the unimodal TPv, 35.6, and the lowest when interacting with the auditory bimodal touchpad TPva, 23.9. An average value of 28.0 was found for the bimodal tactile touchpad TPvt, 24.9 for the trimodal touchpad TPvta and 29.8 for the rotary push-button RPB. A one-factor ANOVA showed significant differences between these HMI variants ($F(4,76) = 4.67$, $p < .01$).

Two interesting aspects of the results require further discussion. First, whereas additional feedback consistently had positive effects, trimodal feedback showed no general advantage over bimodal. The touchpad with combined visual, auditory and tactile feedback did not prove to be the best solution. On the basis of the data available, it is difficult to extrapolate the meaning of this finding for real driving scenarios. It seems very likely that, once a driver listens to loud music or drives on bumpy roads, the disadvantages of one non-visual modality may be compensated for by the advantages of the other. However, this explanation needs experimental tests with real-life driving situations. Second, regarding the bimodal touchpad variants, additional auditory feedback led to somewhat shorter gazes at the display and lower workload ratings than additional tactile feedback. One reason may be that people are more used to integrating auditory feedback into their action strategies than abstract tactile feedback. More practice with Tactons may be required to achieve equal results.

7 Conclusions

The present study has shown that a touchpad can be a reasonable alternative to a conventional rotary push-button switch as a central input device for on-board systems in vehicles. But it also became clear that a touchpad which provides only visual feedback will not be usable while driving. Additional auditory or tactile feedback was necessary to reduce visual distraction, decrease task completion times and reduce the workload associated with the input task. In this regard, the experiment also confirms the expected advantage of multimodal over unimodal feedback. Although auditory feedback yielded somewhat better results than tactile feedback, Tactons proved to

provide a substantial enhancement compared to unimodal visual feedback. In order to give a more detailed answer to the question of the suitability of a touchpad to control in-vehicle applications, a comparison of the multimodally enhanced touchpad with a rotary push-button with multimodal properties, such as BMW's force feedback iDrive controller, would be an interesting next step for future research.

Acknowledgments. We would like to thank Thomas Hempel and Dave Lewis for helpful comments on an earlier draft of this document.

References

1. Norman, D.A.: Psychology of Everyday Things. Basic Books, New York (1988)
2. Gaver, W.A.: Auditory Interfaces. In: Helander, M.G., Landauer, T.K., Prabhu, P.V. (eds.) Handbook of Human-Computer Interaction, pp. 1003–1041. Elsevier, Amsterdam (1997)
3. Wickens, C.D., Hollands, J.G.: Engineering Psychology and Human Performance. Prentice Hall, Upper Saddle River (2000)
4. Wickens, C.D., Sandry, D., Vidulich, M.: Compatibility and Resource Competition between Modalities of Input, Central Processing, and Output. Human Factors 25, 227–248 (1983)
5. Liu, Y.C.: Comparative Study of the Effects of Auditory, Visual and Multimodality Displays on Driver's Performance in Advanced Traveller Information Systems. Ergonomics 44, 425–442 (2001)
6. Cockburn, A., Brewster, S.A.: Multimodal Feedback for the Acquisition of Small Targets. Ergonomics 48, 1129–1150 (2005)
7. Vitense, H.S., Jacko, J.A., Emery, V.K.: Multimodal Feedback: An Assessment of Performance and Mental Workload. Ergonomics 46, 68–87 (2003)
8. Williams, J., Michelitsch, G., van Veen, J.-H., Goronzy, S.: MIAMM: Lessons Learned in Haptic & Tactile Interaction. In: Buss, M., Fritschi, M. (eds.) Proceedings of Eurohaptics 2004, pp. 30–35. Technische Universität München, München (2004)
9. Brown, L.M., Brewster, S.A., Purchase, H.C.: A First Investigation into the Effectiveness of Tactons. In: Proceedings of Worldhaptics, pp. 167–176. IEEE, Pisa, Italy (2005)
10. Brewster, S.A., Brown, L.M.: Non-Visual Information Display Using Tactons. In: Extended Abstracts of ACM CHI 2004, pp. 787–788. ACM Press, New York (2004)
11. Mattes, S.: The Lane-Change-Task as a Tool for Driver Distraction Evaluation. In: Strasser, H., Kluth, K., Rausch, H., Bubb, H. (eds.) Quality of Work and Products in Enterprises of the Future, pp. 57–60. Ergonomia Verlag, Stuttgart (2003)
12. Byers, J.C., Bittner, A.C., Hill, S.G.: Traditional and Raw Task Load Index (TLX) Correlations: Are Paired Comparisons Necessary? In: Mital, A. (ed.) Advances in Industrial Ergonomics and Safety, pp. 481–485. I. Taylor & Francis, London (1989)
13. Diederich, A., Colonius, H.: Bimodal and Trimodal Multisensory Enhancement: Effects of Stimulus Onset and Intensity on Reaction Time. Perception & Psychophysics 66, 1388–1404 (2004)

An European Approach to the Integrated Management of Human Factors in Aircraft Maintenance: Introducing the IMMS*

Marie Ward and Nick McDonald

Aerospace Psychology Research Group, Trinity College, Dublin, Ireland
maward@tcd.ie

Abstract. Previous research in aviation maintenance has highlighted the need to understand normal practice in order to advance the potential impact of Human Factors and bring aviation to a new safety level. What to do with this information then is crucial. What is presented here is an approach to do this, to use this information in such a way that it becomes key to safety and process improvement. This approach is currently being developed within the European funded HILAS project.

Keywords: Aviation Maintenance, Human Factors Lifecycle, Operational Process Model.

1 Introduction

The HILAS (**H**uman **I**ntegration into the **L**ifecycle of **A**viation **S**ystems) project is developing a model of good practice for the integration of Human Factors across the lifecycle of aviation systems. The project contains four parallel strands of work: the integration and management of Human Factors Knowledge; the Flight Operations environment and performance; the evaluation of new Flight Deck Technologies, and the monitoring and assessment of Maintenance operations. This paper reports on work carried out within the Maintenance Strand and the development of an Integrated Maintenance Human Factors Management System (IMMS).

1.1 Understanding Normal Practice

Human Factors has moved beyond analysing human fallibility and related performance deficits. It is increasingly addressing how people behave in normal operational contexts and how performance in such contexts can be better supported by design for use, by better planning and operational management and by quality and safety management systems.

---

* AMPOS (Aircraft Maintenance Procedure Optimisation System), AITRAM (Advanced Integrated Training for Aircraft Maintenance), ADAMS 1&2 (Aircraft Dispatch and Maintenance Safety) and TATEM (Technologies and Techniques for New Maintenance Concepts) are all projects funded by the European Commission.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 852–859, 2007.

Methodologies such as NOTECHS (non-technical skills) [4] and LOSA (Line Operational Safety Audit) have developed out of this awareness. Researchers have gone further and argued that the reasons for the failure of Human Factors research to eliminate what Reason had referred to 10 years earlier as ‘latent pathogens’ was that people were not observant enough to the ‘softer issues’ that were happening in organisations and in particular to management’s ability to manage these softer issues (Westrum, 2001) [8.]. He argued that there was a need to begin to understand the dynamics of organisations at this level, a belief supported by Johnston [3], Pelegrin [5] and Troadec [6]. Recent research has confirmed this view that while Aircraft Maintenance Engineers (AMEs) are at the coal face of aircraft maintenance, and thus often of incident investigations, it is often these ‘softer’ issues and issues that are ‘upstream’ from this coal face that cause the problems and where Human Factors could potentially have a large impact ([7], [1]). Ward’s (2005) research across four aircraft maintenance organisations demonstrated, for example, that the softer issues are what affect people on a day to day basis [7]. Personal relationships and genuineness can facilitate performance and if trust is lacking between people and departments then sharing of information is difficult and work deteriorates.

1.2 Human Factors - A Lifecycle Approach

Understanding these issues is the first step, what happens to that knowledge is the second step. This requires an integrated approach, which systematically generates knowledge about the human aspects of the system at the operational end and transforms this ‘knowledge about’ into an active knowledge resource for more effective management and operational systems and better, more innovative, design. The challenge is to develop and demonstrate a model of Human Factors research, practice and integrated application, linking design and operation – in a ‘system lifecycle approach’. What is being proposed here is such a model of Human Factors.

Human beings are involved in all aspects of organisational life in some way or another. There is thus a lifecycle of application of Human Factor initiatives that follows this involvement throughout the organisation. Any Human Factors model needs to take this into account. However this is something that has been neglected in Human Factors over the years. Often Human Factor initiatives have focused on one or some aspects of this lifecycle, e.g. analysing incidents, creating awareness, making recommendations. Thus a lifecycle approach to the application of Human Factors needs to be taken moving right through the cycle from initial data gathering to implementation, change and innovation. Table 1 presents some of the elements that may go towards representing this lifecycle of the application of Human Factors in the organisation. These are broken into the following categories; task support, organisational learning and human resources, risk management and finally innovation. What needs to be explored is how current initiatives fit in with this lifecycle and then where any possible gaps might exist.

Table 1. Elements of a Human Factors Lifecycle

HF Role - general category	HF Role - lower level category	HF Role - specifics
Task support	Planning	Requirements, Documentation, Customised documentation / Personal notes, Allocation of tasks, How to do it, Overview of task planning
	Task Management	Anticipate critical path, Manage conflicting resources, Communication / information exchange in real time, Co-ordinate with parallel tasks, Management of parts / tools, Current technical data e.g. BITE Data
	Feedback	Sign-off task completion, Handover incomplete tasks, Review and/or report problems, Feed forward potential risk, Task surveys
Organisational Learning & Human Resources	Competence Requirements	Training needs, Initial training, Recurrent training, Embedded / OJT training
	Selection	Task descriptions, Competence Requirements, Assessment Scenarios, Career Framework
	Mentoring	Task and role models, Appraisal and feedback criteria, Learning Objectives, Problem focused support
	Performance Management	Performance indicators for individual and groups, Performance and process outcomes, Evaluation criteria, Establishing performance targets
	Partnership	Common framework for operational analysis (see also risk management), Communication forum, Agreeing programme of action, Implementation and review of action
	Feedback	
Risk Management	Gathering Normal performance data	Performance reports, Suggestions for improvement, Audits
	Gathering non-normal performance data	Incident reporting, Confidential reporting, Quality Discrepancy Reporting
	Managing data	Analysis and diagnosis, Risk assignment, Recommendations
	Decision process	Prioritising actions including CBA to meet competitive requirement
	Planning Change	Deriving requirements for operational change, Deriving requirement for HR development, Achieving consensus, Implementation plan, Developing mgt role to facilitate change
	Managing Implementation	Preparatory and supporting actions e.g. training, Monitoring implementation of decisions, Managing change process to maintain coherence
	Evaluating Change	Evaluation criteria, Reports on implementation progress, Assessment of change
Innovation	Managing Information to Requirements	Sharing Information, Transforming Information, Deriving Requirements
	Implementing Requirements in Design	Transforming HF requirements into system design requirements, Tracking Requirements in Design Process, Evaluation of prototypes / new design
	Certification	Evaluation of new applications of technologies

1.3 Objectives of the HILAS Strand Maintenance Project

The HILAS Maintenance Strand hopes to address some of the areas of the above lifecycle that may not be addressed with current Human Factor initiatives. Many

Human Factor tools and approaches have been developed in previous research projects, for example, AMPOS, AITRAM, ADAMS 1 & 2. HILAS will integrate these tools and add to them if necessary to provide a system that will manage as many of these functions as possible which manage the human component of the system. This system then needs to be evaluated including the effectiveness of the resulting impact on systems, processes and performance. A standardisation process across companies and countries will take place to explore how this system could provide a comprehensive means of compliance with current and future regulatory requirements.

1.4 Organisations Involved in the Research

All of this work is being carried out as part of the HILAS project which runs from 2005-2009 with 39 partner across Europe and China. The partners involved in the Maintenance Strand of the HILAS project are the outlined in Table 2.

Table 2. HILAS Maintenance Research Partners

Industrial Partners	Research Partners	IT Partners
Maintenance Organisations		
Adria Airways, Slovenia	CAUC, China	KITE Solutions, Italy
Atitech, Italy	CERTH / HIT, Greece	
Eurofly, Italy	ICCS, Greece	
SR Technics, Ireland	IFF, Germany	
Manufacturing Organisations	JRC, Italy	
Selex Communications, Italy	NLR, Netherlands	
Thales Avionics, France	TCD, Ireland	
Turbomeca, France		

2 The Integrated Maintenance Human Factors Management System (IMMS)

The following are some key objectives, that were agreed by all Strand partners in advance of a User Forum that took place in September 2006, of the system being developed – the IMMS. These objectives, while proving a challege and setting a large agenda for the IMMS, are what the Users have expressed as their need from the system.

- Improve operational performance; improve the functionality of tasks and processes
- Improve safety performance; reduce Human Factor related risks and increase Human Factor predictability of events
- Improve quality; improve innovation in processes, efficiency and effectiveness of processes
- Improve business performance; cost reduction, recoverability, new means of compliance.

2.1 Initial Design Considerations for the IMMS

It is hoped that the IMMS will work in a number of ways. Firstly through 'front applications' the IMMS will support and gather data from the Aircraft Maintenance Engineers (AMEs), Approved Engineers (AEs) and Inspectors working at the coal face of operations and also from all people engaged in the support of these. Thus there are two types of front application. The first (A1 in the Table 3) is for AMEs and will provide them with improved task support, through a portable handheld device, using new technologies such as Virtual Reality (VR) and Radio Frequency Identification (RFID). Additionally the 'Sensor' functionality of this front application will collect routine information on; task completion times, information sources consulted, track interruptions, etc. automatically. Short digital questionnaires will also be provided to consult AMEs during their work on Human Factor issues like experience, encountered problems etc. Front application A2 is for all of the support functions and will both provide information to them on how to manage the 'softer' aspects of managing the check and any problems encountered. It will also provide a means of allowing personnel to create an 'Alert' if anything is going wrong with the progress of the check to allow for better measures and contingency plans to be taken downstream and alleviate any potential problems that will arise as a result.

Component B (Universal Collector and Processing Engine) of the system will gather data from these two front applications and also data from currently operating systems within the organisation, for example, planning, engineering and quality systems. Component B will allow all of these system to talk to each other and will allow for interrogation and analysis of the data.

All of this data will feed into and continuously update Component C which is the suite of Human Factor Tools and Methods that will manage the human component of the system. These tools as noted above have been developed in previous research projects such as AMPOS, AITRAM, ADAMS 1 & 2 but will be presented here in a complementary suite and a common web environment. A core part of this Component C is the 'Operational Process Model'. Development on this model began in ADAMS2 and is continuing today, being populated by extensive fieldwork in the HILAS project and the TATEM project. Operational systems involve technologies managed by humans (directly or indirectly) to produce some transformation – in manufacture, transport or service. Such systems commonly function in ways which are quite divergent from their specified procedures. They are also quite resistant to change, and those who design such systems often do not understand how they are normally operated. Thus, if we are to be able to better design, manage or improve operational systems, we need to be able to construct models of those systems as they normally function. Such models should include the functional relationships between humans and technologies which enable the system to be activated. Key processes across the line, hangar and OEM operations are being modelled providing data on the normal technical, information and social relational processes of everyday operational life. This model will provide a means of structuring data and will enable the description and diagnosis of the operational system to support requirements for improvement and change in procedures, facilities, technology and human resources.

Finally Component D of the system is concerned with implementation on two levels, firstly on the level of implementing the actual recommendations that come out of the system and secondly with the implementation of system itself. Previous

research has highlighted that it is often at the level of development and implementation of recommendations that Human Factor initiatives fail [7]. There can be an assumption that Human Factor issues are similar to Technical issues and can be dealt with in a similar manner, however this is not the case as usually they are much more complex. Special focus will be given to the development and implementation of recommendations and the tracking and monitoring of the success of the recommendations in dealing with the problem that they were meant to. The latter part of Component D will be dealt with in the next section.

These components of the IMMS are further outlined in Table 3 along with who will be the potential users of each aspect of the system.

Table 3. Components of the IMMS

IMMS	**Front Applications**	**Users**
A1	The 'Front Application' of the IMMS relevant to e.g. AMEs	Aircraft Maintenance Engineers, Approved Engineers, Inspectors Connection to other users: Needs to link with B & C
A2	The 'Front Application' of IMMS relevant to Operational Management and to Support functions, e.g Quality, Engineering, Planning etc.	Users: Operational Management Person power Scheduling and Positioning, HR & Personnel, Customer Relations Manager, Environmental Protection, Training & Learning, Knowledge Mgt, Quality, Operational Planning, Engineering, Safety Management System, HF Analysts, Supply Chain, Workshop, Strategic Planning OEM Users: To be defined Connection to other users: Needs to link with A & C

IMMS	**Back Applications**	
B	Universal Collector and Processing Engine	User – Administrator Connection to other Systems – A&B&D plus Financial systems, HR & personnel systems, Customer Relations Management Systems, Environmental Protection Management Systems, Training & Learning, Knowledge Mgt, Planning, Engineering systems, Safety Management Systems (including Risk Analysis / Management systems), Quality systems (event, incident reports, auditing systems), Supply Chain systems, Workshop Systems, Maintenance Systems - aircraft documentation systems, job performance data, job card system, MPD, Environmental Conditions, Flight systems, Airport Authority Computer Systems, OEM Systems
C	Suite of Human Factor Tools and Methods Operational Process Model	Users: Human Factor Analysts, Business Analysts Connection to other systems: C

IMMS	Organisational Support for Implementation	
D	Organisational Support for the Implementation of recommendations and the IMMS itself. Organisational Learning and Inter-Organisational Sharing of Information and Learning.	Users: Human Factor Analysts, Safety Management System, Quality Dept, Organisational Learning, Training, Business Analysts Connection to other systems: C

3 Organisational Support for Implementation

Component D of the system will also address the wider issue of organisational support. It was felt important to build this into the system and develop 'system requirements' for this as well in order to ensure that it was a priority throughout project development. Sometime issues around implementation can be left to the end of the project which is not sufficient to ensure success.

A key element of any organisational change is the existence of an appropriate organisational culture [2]. The following are necessary in order to evaluate the capacity of an organisational change:

- The theoretical concept of organisational culture should include both concrete and tacit aspects.
- Organisational as well as individual variables must be taken into account.
- The first stage of the culture and change research should include evaluating the organisational capacity to mobilise energy and commitment towards the change implementation.
- The type of change that will be implemented needs to be defined. This information will help to define what organisational and cultural dimensions need to be evaluated and the scope of the proposed implementation culture model. Specifically, two characteristics are important to be clarified: (1) The rate of occurrence; for example: discontinuous, incremental, continuous, etc. (2) More important, the type of change characterised by how it comes about; for example: planned, emergent, contingency or choice.

Also it is important to note that a change or Human Factors initiative often is a chance for people to have a 'voice'. It gives people a chance to say how their work is going for them at the moment [7]. It is thus important in a change initiative to work through some of the existing problems first before embarking on the change initiative or else the initiative can become jeopardised by these existing problems. In order to do this the project is supporting the development and implementation of a systemic approach to learning, innovation and continuous improvement, based on best practice approaches to Organisational Learning enhanced by focused research and development in Human Factors. New knowledge and insights gained will feed into the Operational Process Model continuously creating a viable model of normal practice. This will also contribute to basic and applied Human Factors training.

4 Conclusion and Future Steps

What has been presented here is a high level overview of what we are trying to achieve within the Maintenance Strand of the HILAS project. The overall project has four stands and the objective of looking at the lifecycle of aviation (as opposed to the Human Factors lifecycle). To achieve this, the Maintenance Strand and the Flight Operations Strand have been working in parallel to develop a system that will have applications in both areas and feed forward into new design. The next step involves collaboration with Flight Operations to develop the System Requirements Specification for the elements that are common to both systems and then within Maintenance developing the System Requirements Specification for the other elements. The Knowledge Integration Strand of the project has played a crucial role in enabling this process.

Acknowledgments. The authors would like to thank the HILAS Maintenance Strand members and the European Commission for funding this work.

References

1. Baranzini, D.: Team Systems in Aircraft Maintenance. Doctoral Thesis in press: Trinity College Dublin (In press)
2. Díaz, et al.: Notions about Cultural Implications of Implementation Requirements. HILAS Theoretical Workshop, Tenerife (2006)
3. Johnston, N.: Presentation at the 14th Human Factors Symposium. Organised by Airbus in association with IFALPA. Dublin, pp. 2–4 (October 2001)
4. Lodge, A.E., NOTECHS,: cited. In: Johnston, N. Presentation at the 14th Human Factors Symposium. Organised by Airbus in association with IFALPA. Dublin, pp. 2–4 (October 2001)
5. Pelegrin, C.: Presentation at the 14th Human Factors Symposium. Organised by Airbus in association with IFALPA. Dublin, pp. 2–4 (October 2001)
6. Troadec, J.P.: Presentation at the First European Commission Combined Energy and Transport Conference in Barcelona (October 18-19, 2001)
7. Ward, M.: Contributions to Human Factors from three case studies in aircraft maintenance. Doctoral Thesis, Trinity College Dublin (2005)
8. Westrum, R.: Presentation at the HIAS Conference, Toulouse, France, Eastern Michigan University – Social Scientist (September 2001)

A Research of Speech Signal of Fire Information Display Interface

Liang Zhang[1], Xianghong Sun[1], and Thomas Plocher[2]

[1] State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
[2] Honeywell ACS Labs, Minneapolis MN 55418, USA
zhangl@psych.ac.cn, tom.plocher@honeywell.com

Abstract. This study was conducted to investigate the effect of speech rate and tune on intelligibility of fire information words and sentences under the conditions with different levels of noise. The result showed that the types of signals and noise levels affect the intelligibility significantly .The appropriate tune for fire information display interface is mezzo-soprano. The appropriate voice rate is 5 characters per second for words display, 7 characters per second for usual sentences display and 6 characters per second for the sentences with numbers display.

Keywords: fire information display interface; speech signal; tune; speech rate; noise occultation.

1 Introduction

Fire alarm system is an essential part of high buildings in modern times, which helps firefighters' detection more efficiently and reduces the casualty [1]. The fire information display of the alarm system can show the information inside the building, e.g. build structure, fireplug locations, firing time and locations, etc.

The firefighters have to make decisions depending greatly on huge amount of information provided and reach to their decisions in limited time. It was seen that using the single visual display is likely to result in users' visual overloading, inefficiency of information processing and even mistakes of decision making.

Auditory system is another important sensory system for getting information, which is the major complement to visual system. Moreover, human responds to auditory stimulus is faster than visual stimulus [2]. So using speech signals could be promising to solve the problems mentioned above.

The issue of using speech signals has come to engineering psychologists' attention [6,7]. Intelligibility is the precondition of signal application and the essential evaluating index [3]. The speech rate and tune are the essential properties of speech signals. At present, there have been some related researches abroad [8] but rarely in China. Unfortunately, there are also diversity results among the different language

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 860–866, 2007.

using. And Chinese pronunciation, grammar, expression etc. are all unique from other languages. So the parameters of Chinese speech signals should be studied focused on the property of Chinese.

In addition, the types of phrases, such as the acquaintance to the content, the length of the speech, whether include numbers, etc., also affect the signal intelligibility. There are several related researches (Liu Baoshan, 1995; Zhang Tong, 1997) in China, but all were aviation warning system voice. The signals are verb phrases contained 2-5 characters, without numbers, which are highly professional [4]. Considering the different experimental materials and background, the results, i.e. the speech parameters are likely different. So far, there is not any human factors standardization of speech signals. That's the meaning of this study.

Another important factor considered in this study is the background noise, which has been evidence to affect on auditory signals [5]. The speech signal parameters should vary according to its using situation. Considering the high noisy background in fire emergency, we concerned the noise effect on speech signal, and the appropriate parameter under the high level of noise.

To sum up, this study aims to investigate the speech rate and tune under different levels of noise background, using the fire information phrases. And the result is expected to be helpful for the design of fire information display interface.

2 Method

Design. A 4-factor, within-subjects design was conducted. Each participant experie-nced a total of three experimental sessions. The factors and their levels are as follows.

Table 1. Factors and their levels

Factor	A	B	C	D
	Speech Rate (characters/s)	Tune	Speech type	Noise Level (dB)
Level 1	5	Soprano	Noun	0
Level 2	6	Mezzo-soprano	Numeral	60
Level 3	7	Tenor	General sentence	80
Level 4	8	Baritone	Numeral sentence	
Level 5	9			

Materials. 80 sentences or words, derived from the firefighters' interview and Chinese Fire Fighting Constitutions, are balanced distributed to the treatments combined of A, B, C factors. The sentences' length is 2-4characters, while the words' 6-10 characters.

Index of speech and noise. The speech was monotone without emotion, recorded by t male and female announcers in an isolated booth. The experimental phrases were processed from the records by the sound software, and its time accuracy is +1ms.

The noise background was generated from pink noise. The SNR (Signal Noise Ratio) was 5 dB. The dB size was measured by a SL-4001 sound instrument, the error size of which is +1dB, accurate to 0.1dB.

Participants. The participants in this experiment were 36 males, aged from 20-27 years old. All participants reported well audition and fluent mandarin.

Procedure. There was three experimental sessions, and each session under one noise level. In every session, participants can hear some sentences (or words) from the earphone. There was a pause after each sentence (or word), the participants were asked to repeat loudly what they just heard as far as they can. There were 5 minutes break between two sessions. And the sessions sequence were Latin square balanced.

Data record and process. The participants' respond accuracy and reaction time were collected, and then processed by spss.11.0.

3 Results

There are 34 data available. In the analysis of accuracy, the incorrect, incomplete or vacant reactions were recorded as "0", while the correct reactions were recorded as "1". In the analysis of reaction time, the "0" reaction time was filled with the mean value of that treatment combination.

3.1 Effect of Noise

A repeated-measures statistical analysis was conducted and found the difference of accuracy（F（2，66）=14.04，p<0.001 and the RT（F（2，66）=14.04，p<0.001） are both significant.

The accuracy of 0 dB is significant higher than the other two noise levels. But there is no difference between the 60 and 80 dB. The mean of accuracy is showed in figure1 (the left).

The multiple comparisons on RT showed the differences between any two noise levels. The 80dB is the shortest, then 0 dB, and 60dB is the longest. See figure1 (the right)

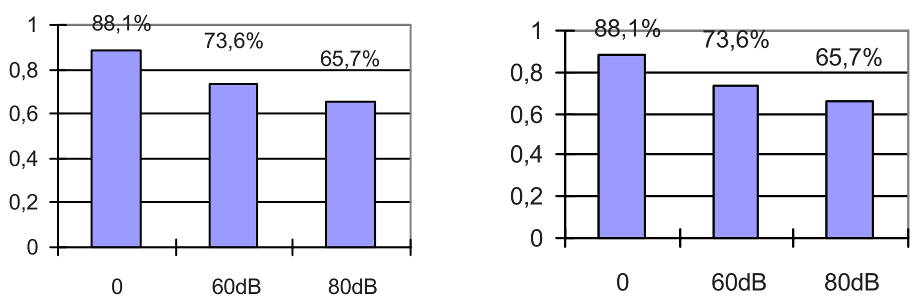


Fig. 1. Means of accuracy (the left) and RT (the right) (ms) under different levels of noise

3.2 Speech Rate

The accuracy descended with the speech rate ascended. Furthermore, the descent velocity varies under different noise. There is a interaction between speech rate and noise level, $F_{accuracy}(8.264)=6.16$, $p<.001$;$F_{RT}(8, 264)=29.07$, $p<.001$. The accuracy descends the most rapidly under the 80dB noise.

Considering the high noisy background in fire emergency, we analysis the data under the high noisy level father more.

Under 80dB, the accuracy of 5 levels of speech rate descend linearly, $F(4, 132)=64.55$, $p<.001$. The difference between any two levels is significant. See figure 2(the left).

The RT varied also, but not linearly, $F(4, 132)=10.975$, $p<.001$. The multiple comparisons showed that the RT was short when 5, 6, 7 characters per second, and there is no significant difference among the three levels. But when the rate quickened up to unclear (the accuracy below 60%), the RT increased sharply. The RT on 8,9 characters per second were longer ($p<.01$) .See figure 2(the right).

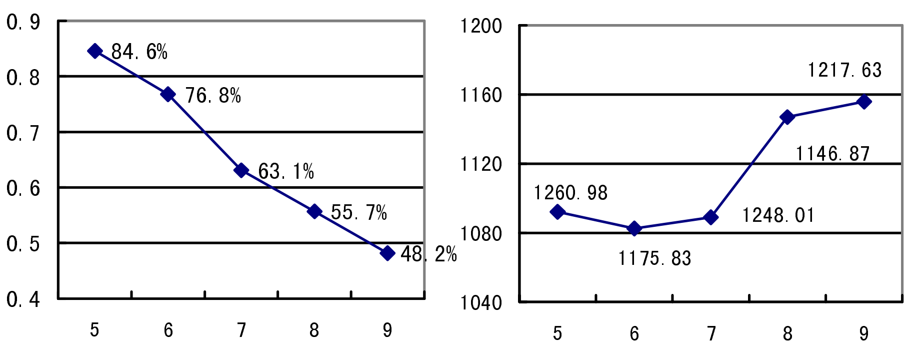


Fig. 2. Means of accuracy (the left) and RT (the right) (ms) on each speech rate (characters/s)

3.3 Tune

The male speech is more likely to be occulted than female's. Moreover, there is an interaction between tune and noise. The difference among tunes is greater under the noisy background and less different when there's no noise. Mezzo-soprano's accuracy is much higher than the others, especially under the 80dB noise. The tune major effect under 80dB is significant, $F_{accuracy}(3,99)=40.245$ $p<.001$; $F_{RT}(3,99)=6.711$,$p<.001$.

The RT of each tune under the 80dB noise level is showed as Table 2. Mezzo-soprano's accuracy is the highest, and tenor and mezzo-soprano's RT are shorter, but there is no statistical difference between them.

Generally, the mezzo-soprano is the optimal for speech signal.

Table 2. RT (ms) of each tune under the 80dB noise level

Tune	Soprano	Mezzo-soprano	Tenor	Baritone
RT(ms)	1138.98±11.35	1101.80±9.78	1093.98±11.35	1118.400±10.41

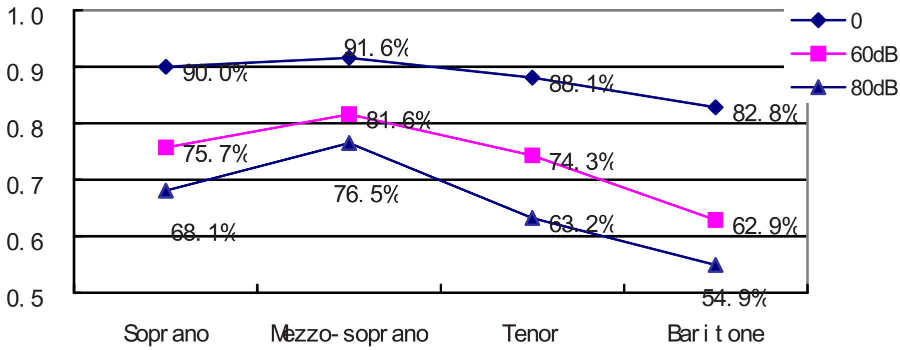


Fig. 3. Accuracy of each tune under the three levels of noise

3.4 Speech Type

The main effect of speech type is significant, $F_{accuracy}(3, 99) = 4.504$, $p<.01$; F_{RT} (3, 99) =226.33. And the interaction between type and speech rate is significant, Faccuracy (12, 396) =17.391, p<.001; F_{RT} (12,396) = 51.37, p<.001. The four types' optimal speech rates are somewhat different.

Farther analysis showed that: (a) looking at the aspect of speech length, the sentences' accuracy descend faster than words', while their RT ascending is similar; (b) looking at the aspect of the content, signals containing numbers descend faster than those without numbers, and the RT ascending faster also. Considering the factors above, the four types' appropriate speech rate are as follows:

Table 3. The optimal rate(characters/s) of each speech signal type

Type of speech signal	Noun	Numeral	General Sentence	Numeral Sentence
Speech rate (characters/s)	5	5	7	6

4 Discussion

4.1 Effect of Noise

In the researches of display design, it's very important to conform the experimental condition to the use situation. Therefore, the design of speech signal on fire display shouldn't ignore the high noisy background.

It's obvious that the noise occultation affect the speech signals intelligibility. The intelligibility descended with the noise heightened. That means the speech rate should be relatively slowed in the noisy environment.

Further more, the noise occultation is close related to the voice properties, such as pitch or audio-frequency etc. According to the previous research, basso is prone to be occulted. That's the reason why we didn't make the bass or alto as an experimental level. The results also showed female signals are superior to males', and the superiority comes greater in noisy environment.

Besides, we compared the high noisy level with the low noisy and without noise level. The different results among the three conditions may indicate a general trend of noise effect to speech signal, though the details still need more researches.

4.2 Acquaintance Effect to Intelligibility

Contrastively, the accuracy in current study is inferior to some researches before. That can be due to the unfamiliarity to experimental materials. The experimental phrases are highly professional in the previous researches, so participants were asked to learn the phrases beforehand. But the specialization of this experimental phrase is much lower. Thus we didn't request participants to learn the phrases. That's the major reason that leads to the inferior intelligibility.

Reviewing two related researches, we found the acquaintance effect also. Comparing Liu Baoshan's (1995) study with Zhang Tong's (1997), the former accuracy was higher then the latter, though their materials were both fighter warning system voice. It is mostly because the former participants were pilots while the latter were undergraduates.

Moreover, we found the practice effect in our research. It suggested the acquaintance to signals could affect their intelligibility. So the further study plans to use firefighters as participants. And the better intelligibility is anticipated because the firefighters are much familiar with the fire phrases. The relationship between acquaintance and signal intelligibility can be research farther by comparing the current results with the firefighters' research.

4.3 Contrast

Table 4 lists the contrast among the results in current study, the criterion of American military systems design and the voice parameters in Chinese fighter warning system.

As mentioned above, there isn't a uniform criterion for speech signal design in China. The speech signal not only varies with the languages, but also varies with the speech properties (e.g. content, length, and receiver's attention state etc.).

Compare with the warning voice, the speech signal on display interface is more complex. It can express more information, so its content and length are more various. Moreover, there is a distinction between them. That is the receiver's attentive state. The warning voice often broadcast without any preparation. So it requests to attract receivers' attention rapidly and forcefully. On the contrary, receivers usual concerned with the display when they use it, so they would be easier to receive the speech signal information. That's why the signal speech rate is faster than the warning voice.

Table 4. Contrast with other parameters of auditory signal

	Type of Task	Tune	Voice rate	RT(s)
American voice warning		Female	156~178 words/min	
Chinese voice warning In fighter cockpit	Single-task	Mezzo-soprano	180~300 ch/min	
	Double-task		240~360 ch/min	2.181
Speech signal in fire interface	Single-task	Mezzo-soprano	300~420 ch/min	1.102

4.4 Further Research

This study simply concentrated on the auditory signal design. In fact, multiple display interfaces is more efficient than single visual or auditory interface. Therefore, further researches plan to explore the combination of visual and auditory interface. The multiple interfaces will be more efficient and humanized than single interface, and its application is anticipant to be exciting.

5 Conclusion

The result showed that the types of signals and noise levels affect the intelligibility significantly. Under the high noisy background, the optimal tune for fire information display interface is mezzo-soprano. The optima speech rate is 5 characters per second for words display, 7 characters per second for usual sentences display and 6 characters per second for the sentences with numbers display.

References

1. Zhijian, F.: Development of human-machine interaction:multimedia and multisensory. Human Factors 2, 34–38 (1998)
2. Quan, P.: Design of application Interface based on the human cognition. Computer engineer and application 19, 148–150 (2001)
3. Tong, Z.: The research of voice rate for warning system. Chinese Journal of Applied Psychology 1, 34–39 (1997)
4. Xioxia, G.: Selection of Chinese phrase for fighter voice warning system. Chinese Journal of Aviation Medicine 5, 230–231 (1994)
5. Baoshan, L.: Determination of design parameters of synthesized Chinese voice speech for auditory warning system in fighter cockpit. Human Factors 1, 24–27 (1995)
6. Minnesota, S.J., California, W.A.: Critical issues in the development of speech technology for manmachine
7. Integration. In: Proceedings of 26th Annual Meeting of the Human Factor Society (1982)
8. Steeneken, H.J.M., Langhout, G.: Voice interactive system in severe noise condition, (1985) ADP005603
9. Simpson, C.A.: Synthesized speech rate and pitch effects on intelligibility of warning messages for pilots. Human Factor 5, 32–33 (1984)
10. MIL-STD-1472D. Human engineering design criteria for military systems. Equipment and facilities, pp. 39–46 (1989)

HCI Testing in Flight Simulator: Set Up and Crew Briefing Procedures. Design and Test Cycles for the Future

Rolf Zon and Mariska Roerdink

National Aerospace Laboratory NLR, A. Fokkerweg 2, 1059 CM Amsterdam, Netherlands
zon@nlr.nl, mroerdin@nlr.nl

Abstract. Within the HILAS flight deck strand a flight simulator experiment was performed. It was the first experiment of a series of two. The aim of the experiment was to evaluate a set of Human Factors tools (toolbox) for measurement of HCI aspects when new technologies are installed on a flight deck. This paper is an informative document that describes the tools and technologies that were applied in that first experiment. The lessons that will be identified from this experiment will be the input for a better and improved Human Factors toolbox which will be evaluated in the second experiment. Eventually this toolbox may be used by authorities and industry as a structured way of measuring HCI aspects of new technologies and applications that may be installed in future flight decks and as such may also be used as a HF certification instrument.

Keywords: HILAS, flight deck, human factors, experiment, flight simulation, HF toolbox, certification.

1 Introduction

1.1 The HILAS Project

HILAS stands for Human Integration into the Lifecycle of Aviation Systems. The general aim of the project will be explained in other papers at HCI International 2007: McDonald (2007), Jorna (2007), Ward and McDonald (2007), Cahill et all (2007).

The flight simulator experiment that is discussed in this paper is part of the flight deck technologies strand[1] of HILAS, which is aimed at creating a Human factors related 'toolbox' that can be used to design and evaluate (new) flight deck technologies and their applications.

---

[1] The HILAS project runs from June 2005 until June 2009 and is funded by the European Communities as part of the 6th framework. The project partners in the Flight Deck Strand are: Smiths Aerospace and BAE SYSTEMS from the UK, NLR, TNO, Noldus and the University of Groningen from the Netherlands, Elbit from Israel, Galileo Avionica and Deep Blue from Italy, STL from Ireland and Avitronics from Greece.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 867–876, 2007.

1.2 The Experiment

Within the flight deck strand two major experiments are scheduled. The first one will be discussed in the current paper. The lessons identified from the first experiment will be used to improve the tools, technologies and applications for the second, and last, experiment within the HILAS flight deck strand. This is a unique approach for this kind of research. The same group of partners designs the first experiment and learns to collaborate more effectively. The second experiments will validate if the improvements made, were indeed resulting in better human performance and safety.

The primary aim of the experiment was to evaluate possible future technologies by means of the HF toolbox. The researchers wanted to find out if important tools or methodologies are missing. If the latter is the case, these missing tools and methods need to be identified and added for the second experiment. A second aim of the experiment is to identify suggestions for improving the technologies and applications that were installed in the cockpit. As such the HF toolbox as well as the technologies and applications will be improved when they are used in the second experiment.

One of the major reasons for designing such an HF toolbox is that can be applied as an instrument for HF certification. The importance of HF certification and the added value of having a toolbox to aid certification is explained by Jorna (2007).

1.3 Flight Simulator

NLR's Generic Research Aircraft Cockpit Environment, GRACE, is a generic flight simulator, representing a modern large two-engine fly-by-wire airliner. For the current experiment a generic Airbus configuration was selected. A high fidelity simulator like GRACE allows researchers to perform realistic experiments in a fully controlled environment, but including the opportunity of installing new equipment, like the set of HILAS technologies in the cockpit.

The generic cockpit features in total six Liquid Crystal Displays (LCD). The display formats are equivalent to existing Airbus A320 cockpit displays. Examples of the Primary Flight Display, the Navigation Display, the Engine Display and the System display are shown below.

Fig. 1. GRACE (Generic Research Aircraft Cockpit Environment)

Fig. 2. The cockpit displays in GRACE in Airbus configuration

More information about GRACE may be found in Egter van Wissekerke, R.F. (2004).

2 Tools

The toolbox comprised tools that are applied during the design phases and/or the evaluation phase and for analysis of results. These tools range from design for display formats (VINCENT) to software for analysing eye scanning patterns (the GazeProc software package) or psycho-physiological variables (HEART and CARSPAN) until software that recognises complex (behavioural) patterns (Theme) of pilots.

A selection of the tools used during the experiment is described in the next section.

2.1 Eye Tracking

Eye tracking, brought into the experiment by the NLR, is considered to be an indirect measurement of pilot attention and focus on tasks. The assumption is that looking at a location represents that attention is focussed. But looking is not always 'seeing'. You could be thinking about something else. Therefore, eye tracking is always combined with a measurement indicating the actual processing of information, such as heart rate changes or pupil size alterations (processing is associated with increased size) The eye tracking software calculates fixations, dwell time of both pilots on different displays and panels in the cockpit as well as eye blink frequency as an additional indicator of visual workload.

Fig. 3. The ASL eye tracking system

2.2 Psycho-Physiology

Heart rate, respiration rate and blood pressure Pulse Transit Time (PTT) were brought into the project by the University of Groningen and the NLR. They were recorded on the Vitaport system. From the heart beat recordings, the heart rate and heart rate variability could be calculated with the HEART and/ or CARSPAN software packages. Psycho-physiological variables were recorded as indicators of pilot mental workload, either phasic changes to information changes or tonic work load

Fig. 4. Heart rate signal, electrode for heart rate recording and sensor for PTT measurement

2.3 Gwylio

Gwylio, brought into the project by BAE SYSTEMS, is a tool that asks the pilots questions during the flight. This is the only tool which interferes with the pilots

operations during the flight. It asks two kinds of questions, workload ratings on a five point scale and multiple choice questions. The questions were presented on a HMD or on a tablet PC. As soon as a question popped up, the pilots were asked to give a rating on a dedicated device. The aim of applying this tool was to find out to what extent an intrusive way of asking questions influences pilot performance.

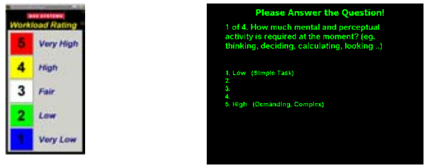


Fig. 5. The two types of questions that were offered to the pilots by the Gwylio tool.

2.4 Electronic Questionnaires

Series of open and multiple choice questions were asked after each flight. They could be answered on an electronic flight bag (EFB) in the cockpit (see photograph below). The Situation Awareness Rating Technique (SART) and the Rating Scale Mental Effort (RSME) were also presented via the EFBs.

Fig. 6. EFB in GRACE

The questions were created by all partners and brought together into questionnaires by the University of Groningen and the NLR. The EFBs came from the NLR.

3 Simulator Data

NLR's GRACE recorded flight parameters as well as pilots controls and switches inputs. Other (subjective) indicators of pilot performance are the video recordings (including audio) that were taken in the cockpit. These will afterwards be reviewed by experts who will give a rating for the pilot performance, assess their situational awareness etc.

3.1 CRIA

Is an interview technique that stems from the social sciences and is brought into the project by Deep Blue. It was originally used in Air Traffic Control (ATC) simulator experiments. In the current experiment it is studied whether this technique is beneficial for flight simulator experiments as well.

3.2 IR Camera

The IR camera observing the pilot, that was provided by TNO allowed measuring the temperatures of the pilots face. The reasoning is that when mental workload increases

respiration increases accordingly. Also, the blood flow to the brain is increased. As such the temperature of the nose should decrease when mental workload increased. Whether that applies in a cockpit setting is part of the study.

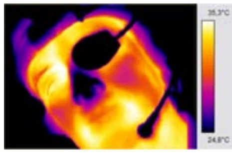


Fig. 7. IR photograph of a pilot wearing an eye tracker in GRACE

4 Technologies and Applications

During the experiment the pilots were confronted with a set of new technologies that were *experimental* yet, but that were developed to such an extent that they could be implemented in GRACE. Such a set of technologies was perfect for evaluation of the HF toolbox since this set of technologies definitely allowed the toolbox to indicate potential design issues. Virtual testing allows learning the lessons before the accident.

4.1 Cockpit Displays

In addition to the standard Airbus cockpit displays, for the HILAS experiment a number of new, not yet in normal Airbuses existing, displays were installed. They will be described below.

4.1.1 Head Mounted Display

The Head System’s hardware was designed by Elbit while the symbology and software running on it was designed by NLR (see photographs below).

Fig. 8. The head mounted display and an example of the symbology of the HMD

The Head Mounted Display provides the pilot with symbols overlaid on the scenery.

- Some of the symbols are stationary and do not move relative to the pilot's head (e.g. air speed).
- Other symbols depend on the pilot's head direction and are actually static relative to the scenery (e.g. navigation marks).

4.1.2 Dual Layer Display

The dual layer display, which was brought into the project by TNO and on which NLR’s software was running, is a display with two visible display layers: one presented in front and the other in the rear or background. The depth-gap between the two layers is 15 mm. The PFD was presented on the front layer and the synthetic terrain + tunnel in the sky (TIS) on the rear. In the baseline experimental condition all information was shown on just a single layer (the rear).

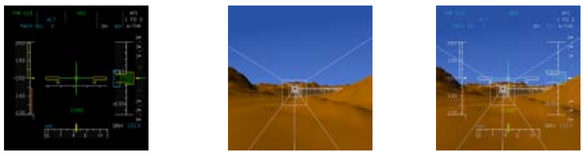


Fig. 9. The PFD (front) + terrain and TIS (rear) = combined on the dual layer display

The application presented on the dual layer display is a 3D synthetic vision application. On this display, an additional dimension is presented to a 2D display, which is the distance along the flight-path. The pilots can use this instrument for guidance to manually conduct landings, even in poor outside visibility.

On the display, the trajectory that should be followed is visualized by means of a Tunnel-in-the-Sky (TIS). A TIS consists of subsequent tunnel frames that are connected at the frame corners. To follow the specified trajectory, the pilot should control the aircraft to fly through the centre of each tunnel section.

4.1.3 Moving Map Display

A moving map system, which was designed by BAE SYSTEMS, comprises an electronic geographical information display combined with location information giving the current position of the aircraft. The electronic map can present recent and 'live' information and the view can be adjusted by the pilot, and some calculations can be automated. Geographical information can be provided in 'layers' which can be selected to present the information required for a particular task, or removed to minimise clutter and distracting information. An example of the moving map can be seen below.

Fig. 10. Examples of screenshots of the moving map

Method of Operation

- The moving map was integrated into the GRACE simulator.
- The aeroplane’s position was automatically displayed on the moving map.

4.1.4 Input Devices

There were two new input devices introduced in the simulator. Both are developed by Smiths Aerospace and both are dedicated technologies for making radio frequency changes. They are described below.

4.1.5 Direct Voice Input (DVI) System

The speech interface to the flight deck is a user independent software based recogniser. To operate the speech recognition system the pilot should press the Push-To-Talk (PTT) button which is located outbound on the stick (next to the autopilot disconnect button, see Figure below), wait for about 1 second and then issue the required syntax. Once finished, the pilot should pause (about 1 second again) before releasing the PTT. The system will then acknowledge the input and carry out the command.

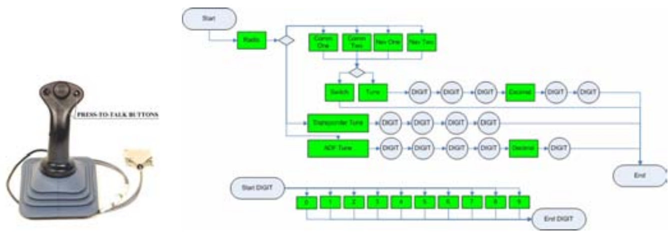


Fig. 11. The side stick used in the HILAS experiment. Notice that there is an additional button (compared to the Airbus side stick) to activate the DVI system. Plus scheme describing the syntax that the pilots could use to make radio frequency changes.

The functionality that is controlled by the DVI provides visual feedback in the form of system / display changes on the radio panel. The pilot could see that the system has correctly implemented his / her command. In addition, auditory feedback was given in the form of a repetition of what the system recognised.

4.1.6 Interseat Touch Screen

Located in the Interseat Console is a full colour, touch sensitive screen, Graphical Input Device (GID, see below), which replaces the current radio panel.

Fig. 12. Screenshot of GID

When the pilot selects a radio, he / she can carry out the following actions:

- Toggle audio output for that radio
- Control the volume
- Enter a new frequency, either from a list or by entering a frequency
- Swapping the active frequency with a standby frequency.

4.1.7 Adaptive Automation

A first implementation of adaptive automation (aiding the pilot based upon an assessment of the pilot mental workload and tasks that need to be performed) was made. During the experiment a number of simulator parameters were assessed in real time as well as the flight phase and tasks that the pilots were working on. In future updates of this adaptive automation program eye tracking data (for assessment of which pilot is doing what) and heart rate and pupil size (for assessment of pilot mental workload) will be included. Eventually such information will be used for giving pilots help or advice, regarding the tasks that they are working on, or less crucial information will be held back until the crew has got less workload, and as such more time, to pay attention to it (see Jorna 2007).

5 Running the Experiment

5.1 Experiment Preparation

Preparation requires time and a strict order of operation. That is especially true with so many partners, technologies, applications and tools. To give an impression of the complexity of how all of these fit together in one big experiment the scheme below is added. It comprises almost all of the technologies, applications and tools with arrows indicating how they are all interconnected.

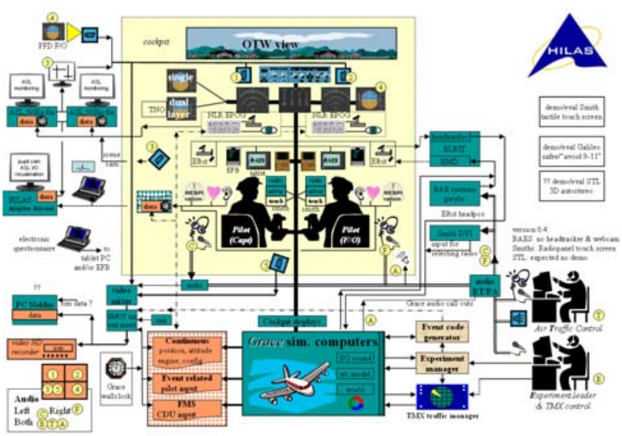


Fig. 13. An overview of how all technologies and tools were interconnected in the experiment

5.2 Training and Preparation

To enable pilots to participate and understand how to operate all the new technologies and the specific characteristics of the simulator a great deal of training is needed. That was accomplished by a briefing guide that was send to the pilots in advance. Furthermore they received a briefing at NLR in which details were explained and with plenty of opportunities to ask questions. After that the pilots made a number of flights in the simulator. After the first flights the different technologies and applications were started so that the pilots could familiarise with those as well.

5.3 Design

A within subjects and between flights experimental design was used. Sixteen airline pilots participated in the experiment. The crews came to the NLR for two days to get a training / familiarisation, perform the flights and discuss with the researchers about their experiences. Pilots always evaluated one technology at the time and there always was a baseline flight that served as reference against which the results of the flights in which a new tool was introduced could be compared. The experiment was split into four blocks. In each block different technologies were evaluated so that pilots would never have to focus on more than one technology per pilot per flight. As such there were twelve flights per crew. All flights or segments took between 10 and 20 minutes. In order to make the experiment more realistic the pilots received cockpit documentation like checklists, load sheets, maps, etc.

6 Closure

At the time that this paper was written the experiment was just completed and the data analysis has just started. Therefore the paper focuses on what was implemented in the experiment. The HILAS project runs until June 2009. The results and conclusions of the analysis of the current and the next flight deck strand experiments will be published in the coming months in papers, articles and reports.

Acknowledgements. This paper was prepared with information from the HILAS flight deck strand project partners.

References

1. Cahill, J., Mc Donald, N., Ulfvengren, P., Young, F., Ramos, Y., Losa, G.: HILAS Flight Operations Research: Development of Risk/Safety Management. Process Improvement & Task Support Tools. HCI International (2007)
2. Egter van Wissekerke, R.F.: GRACE - Generic Research Aircraft Cockpit Environment. NLR-Memorandum ATTH-2004-014, Amsterdam (2004)
3. Jorna, P.G.A.M.: Human performance enhancements: from certification to HCI innovation. HCI International (2007)
4. McDonald, N.: Human Integration in the Lifecycle of Aviation Systems. HCI International (2007)

5. Human Factors Working Group for U.S. Department of Transportation Federal Aviation Administration, Advisory Circular (2004)
6. Roerdink, M.I., Zon, G.D.R.: Briefing guide GRACE experiment January - February 2007, NLR CR-2006-806, NLR, Amsterdam (2006)
7. Ward, M., McDonald, N.: An European Approach to the Integrated Management of Human Factors in Aircraft Maintenance: Introducing the IMMS. HCI International (2007)

Author Index